AN APPLICATION OF BIMOLECULAR FLUORESCENCE COMPLEMENTATION (BIFC) FOR THE DETECTION AND ANALYSIS OF PROTEIN INTERACTIONS ALONG THE ESCHERICHIA COLI TWIN ARGININE TRANSLOCATION (TAT) PATHWAY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jan Stefan Stanisław Kostecki

August 2009

AN APPLICATION OF BIMOLECULAR FLUORESCENCE COMPLEMENTATION (BIFC) FOR THE DETECTION AND ANALYSIS OF PROTEIN INTERACTIONS ALONG THE ESCHERICHIA COLI TWIN ARGININE TRANSLOCATION (TAT) PATHWAY

Jan Stefan Stanisław Kostecki, Ph. D.

Cornell University 2009

The twin arginine translocation (Tat) pathway of *Escherichia coli* possesses an innate ability to translocate fully folded proteins across the bacterial inner membrane; however an *in vivo* method to directly monitor the protein interactions involved in this pathway did not exist. By using yellow fluorescent protein bimolecular fluorescence complementation (YFP-BiFC), protein-protein interactions can now be visualized in unprecedented clarity and at near real time rates along the entirety of the Tat pathway.

Two interacting proteins previously identified and characterized in the Tat pathway, DmsA and DmsD were chosen for YFP-BiFC proof of concept studies. Protein fusion chimeras were created, whereby YFP was split into two fragments, Y1 and Y2, and then attached to the C-terminus of DmsA and DmsD, respectively. Upon coexpression of the two chimeric proteins *in vivo*, DmsA and DmsD interacted, YFP was reconstituted, and upon excitation resulted in the emission of a fluorescent signal.

To demonstrate the utility of YFP-BiFC beyond DmsA and DmsD, we made protein chimeras targeting every part of the Tat pathway. With these chimeras, we were able to detect a fluorescent signal for interactions between substrate-chaperone, substrate-machinery, chaperone-machinery, and machinery-machinery interactions. From these interactions, a quantitative fluorescent signal was obtained, showing a dynamic range in signal intensity depending on the type of interaction being

monitored. Additionally, *in vivo* localization of the protein chimeras could be determined by fluorescence microscopy.

Furthermore, we expanded the applicability of YFP-BiFC in four ways; 1) we generated a DmsD library to isolate higher affinity DmsA binding variants by screening for an increase in YFP-BiFC signal, 2) we used the irreversible association of YFP-BiFC to purify the DmsA and DmsD complex for crystallography studies, 3) we used a family of *de novo* designed 3-helix bundle proteins to investigate the ability of the Tat pathway to interact with proteins of varying degrees of stability, and 4) we were able to obtain a FRET signal between the DmsA-DmsD YFP-BiFC complex and TatC-CFP.

Overall, YFP-BiFC is a powerful tool for monitoring protein interactions *in vivo* and as a stabilizing force for *in vitro* protein analyses.

BIOGRAPHICAL SKETCH

Jan Stefan Stanisław Kostecki was the first of three children born to Zbigniew and Małgorzata Kostecki in Riverside, California. Being the son of an engineer and a pharmacist, his upbringing resulted in a blended understanding of the applications of science and the complexity of biological systems. His academic accomplishments resulted in his acceptance to the University of California, Berkeley where he obtained a Bachelors of Science in Bioengineering in 2004. After Berkeley he started his Ph.D. at Cornell University in the Biological and Biomedical Sciences Program in the Field of Pharmacology, but his engineering background brought him back to the College of Engineering in the Department of Biomedical Engineering. He subsequently joined the laboratory of Dr. Matthew P. DeLisa and received his Doctorate of Philosophy in August 2009.

Byliście mi przykładem

Wszystkiego co się nauczyłem

Wszystko co ja zrobiłem

Wszystkim kim ja jestem

I wszystko co mi przyszłość przyniesie

Będę to zawsze pamiętał.

Wasz przykład dał mi odwagę do zdobycia i podbicia świata.

Dedykuję moim rodzicom tą pracę doktorską

Zbigniewowi i Małgorzacie Kostecki

ACKNOWLEDGMENTS

First and foremost, to my adviser Dr. Matthew P. DeLisa, who four years ago brought me into the DeLisa Research Group, and who saw the potential that I could contribute even before I ran my first experiment. Secondly, to the Cornell University Biomedical Engineering Department that funded three semesters of my teaching assistantship position in BME – 301. Thirdly, I'd like to thank the following people who have made my time at Cornell priceless. Belinda Floyd for everything she has done, BME related or not, her open door policy for having time to chat (no matter how busy she was), and for becoming my "Mom" in Ithaca during my Ph.D. pursuit. To Christian Feiler, for enlightening me with a crystallographers knowledge regarding the magical secrets required for protein purification. To the all the BME – 301 teaching assistants: Frank Kung, Rebecca Moy, Sungkwon Kang, Jan Ma, and Regina Salvat. To Bina Lokchander, whose friendship has continued since graduate school year 2 onward, who spent countless hours with me in Kimball hall, preparing for laboratory experiment days, grading homework, and keeping me sane that first semester of BME – 301. To the Cornell Ballroom DanceSport Team, for the countless friendships made and the endless hours spent on the ballroom floor that provided a much needed balance to the daily research grind. To Becky Graves and Ellan Spero who both taught me how to forget the world for 90-second intervals while the music was playing. To my ballroom partners, Elizabeth Jordan, Rachel Matusow, Margareth Tran, Alissa Mrazek and Jennifer Burt for having our practice pay off. To my motorcycle buddies Kusal De Silva and Filipp Akopyan for those long rides during the warm Ithaca summer.

Finally, thank you to everyone who was not mentioned above. Without you, my experiences in Ithaca, at Cornell University, and the pursuit of this Ph.D would not have been memorable.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa	Amino Acids
ANS	8-anilino-naphthalene sulfonic acid
BiFC	Bimolecular Fluorescence Complementation
CFP	Cyan Fluorescent Protein
D	Aspartate
DsRED	Discosoma sp. red fluorescent protein
E. coli	*Escherichia coli*
F	Phenylalanine
FACS	Fluorescence Activated Cell Sorting
FLAG	A epitope tag for Western Blotting (DYKDDDDK)
FRET	Fluorescence Resonance Energy Transfer
H	Histidine
HIS	A epitope tag with multiple histidine residues (HHHHHH)
HRP	Horseradish peroxidase
I	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Lysine
kDa	kilo-Dalton
L	Leucine
mRFP	Monomeric Red Fluorescent Protein
nm	nano-meter
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCA	Protein Fragment Complementation Assay
PDB	Protein Data Base (http://www.rcsb.org)
POI	Protein of Interest
pmf	Proton motive force
R	Arginine
R1	mRFP1 fragment coding for amino acids 1-154
R2	mRFP1 fragment coding for amino acids 155-238
S	Serine
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SsrA	A protein degradation tag
T	Threonine
Tat	Twin Arginine Translocation
V	Valine
wt	Wild Type
Y	Tyrosine
YFP	Yellow Fluorescent Protein
Y1	YFP fragment coding for amino acids 1-157
Y2	YFP fragment coding for amino acids 158-238

PREFACE

This Ph.D. dissertation has been written as two parts, whereby the first part (Chapter 1) demonstrates in fine detail the *in vivo* ability of YFP-BiFC to detect a variety of protein-protein interactions (e.g. cytoplasmic, transmembrane, and cytoplasmic-transmembrane) throughout the *E. coli* Tat-pathway. The second part (Chapters 2 – 5), demonstrates how YFP-BiFC can be expanded and applied to library screens, in protein purification and crystallography studies, in understanding how *de novo* designed proteins interact with Tat components, and how multiple protein-protein interactions can be detected through a combined YFP-BiFC-FRET analysis.

It is the author's intent that readers gain an understanding of the methodology of YFP-BiFC, view examples of how and where YFP-BiFC has been applied towards the understanding of protein-protein interactions in the Tat pathway of *E. coli*, and subsequently utilize this dissertation as a catalyst to develop novel applications of this technique to further elucidate scientific inquiries.

CHAPTER 1

THE TWIN ARGININE TRANSLOCATION (TAT) PATHWAY – MONITORING PROTEIN INTERACTIONS VIA YFP-BiFC

1.1 Introduction

The Twin Arginine Translocation (Tat) pathway was initially discovered in 1991 by researchers studying protein export in chloroplasts (7,8). Tat directed protein secretion is unlike the well studied general secretory (Sec) export pathway (9-12); notably for the ability of Tat to export fully folded proteins across energy transducing membranes (13). Furthermore unlike Sec, Tat transport has three distinct characteristics: 1) the energy required to move proteins across membranes is derived from the proton motive force (pmf) existing across charged membranes, and not from the hydrolysis of adenosine triphosphate (ATP) as in the Sec pathway (14,15)) the signal peptide that targets the proteins to the Tat pathway is longer than signal peptides in Sec, has a predicted alpha-helical secondary structure (16), and contains two highly conserved arginine residues; 3) the Tat pathway is not a universally conserved protein export pathway (17,18) but to date has been identified in archaea, bacteria (both Gram-positive and negative), chloroplasts, and plant mitochondria (19).

Finally, the Tat pathway does not demonstrate any homologous translocation machinery components in eukaryotic organisms. Elimination of Tat pathway components results in the attenuation of virulence in numerous bacterial pathogens (20-24). From these results, the development of Tat-pathway specific bactericidal or bacteriostatic drugs has been postulated (25), however one of the major limiting factors in the development of these therapeutics is the creation of facile dynamic reporter assays that can monitor protein interactions through the Tat pathway. A proof

of concept reporter assay that confers antibiotic resistance via protein fragment complementation when the Tat pathway is functional has been demonstrated in *Mycobacterium tuberculosis* in an academic setting (26), however converting this assay for industrial scale-high throughput drug screens has yet to be attempted (27).

Here we present a fluorescent reporter with a wide dynamic range that can be used *in vivo* to monitor the protein interactions across the entirety of the *E. coli* Tat pathway and which has the potential for high throughput analysis.

1.2 The E. coli Twin Arginine Translocation (Tat) Pathway

The *E. coli* Tat pathway consists of two distinct protein components that function synergistically with one another and allow for the export of proteins (also known as substrates) across the bacterial inner membrane. The first protein component, the Tat translocation complex, consists of three integral membrane proteins TatA, TatB, and TatC (Figure 1.1 A). In *E. coli* TatA and TatB are single-span transmembrane proteins that share a similar structure and sequence, however TatB has evolved and diverged in sequence from TatA (18), while TatC contains six transmembrane spans (28). Additionally in *E. coli* TatA has been duplicated to give rise to TatE, however TatE is expressed at a low level and deletion of this protein does not affect export of Tat substrates into the periplasm (29). If any of the proteins in translocation complex are deleted (with the exception of TatE), export of Tat-specific proteins will be inhibited, the proteins will accumulate in the cytoplasm, and the cell phenotype will switch from single rod-like shaped cells to elongated filamentous chains due to the cell wall amidase *lpxC* that is not exported into the periplasm by the Tat pathway (30). Truncation analysis (31), site-directed mutations (32-39), and domain-swapping (31) of the translocation complex proteins have been evaluated and have allowed for a better understanding of how the translocon functions and how Tat

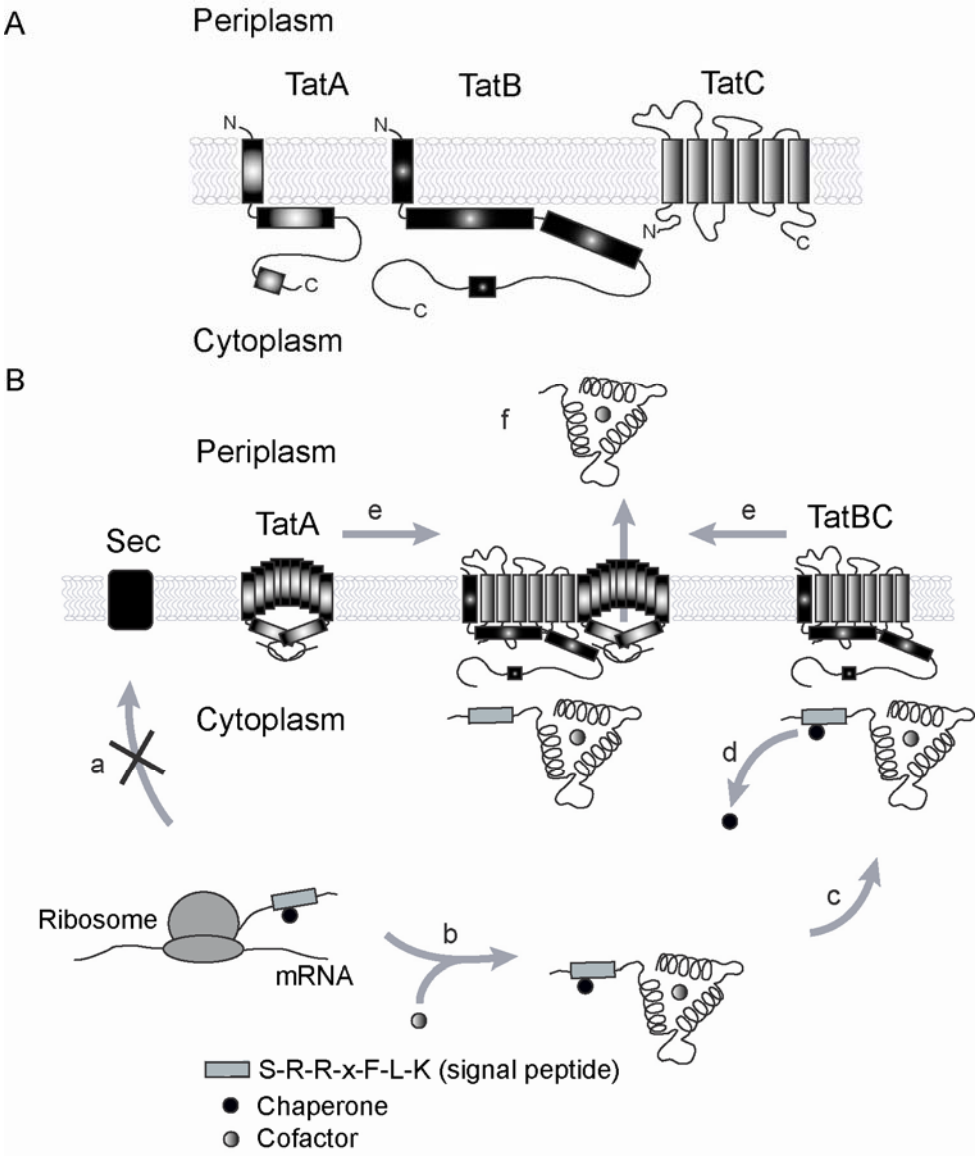

Figure 1.1 Graphical depiction of the Tat Pathway machinery components and timeline of export through the Tat pathway. (A) The three transmembrane components that make up the Tat Pathway. (B) DNA is transcribed to mRNA and is bound by a ribosome that initiates synthesis of a protein, exposing the signal peptide. The signal peptide is bound by a Tat-chaperone and inhibited from Sec export (a). After protein release from the ribosome, tertiary protein structure is attained in the cytoplasm and cofactors are incorporated (b). The protein-chaperone complex is directed to the TatBC complex (c), the chaperone is released, TatA is recruited (e), creating the TatABC translocon. The protein is exported into the periplasm, the signal peptide is cleaved (f), resulting in a mature protein. Adapted from Lee *et. al.* 2006 (5).

substrates interact with the translocon and are subsequently transported into the periplasm.

These membrane proteins have been observed to form two distinct complexes (Figure 1.1 B): one that is comprised of multiple subunits of TatA and a second that contains predominantly TatB and TatC proteins (40-42). TatA homo-oligomers have been observed to form a variable diameter ring structure that may serve as a protein-conducting channel (43) or a patch that facilitates translocation by local destabilization of the bilayer (44). TatB when expressed in *E. coli* strains without other Tat machinery components can self associate and form homo-oligomeric complexes (5,45), however when TatC is present this homo-oligomeric complex is disrupted, in favor of the heteromeric associated TatB-TatC proteins that form a complex to which substrates initially bind (46), suggesting that TatBC serves as the twin-arginine signal peptide binding site.

The Tat-pathway proteins and their associated chaperones, make up the second protein component of the Tat pathway. Proteins destined for the Tat pathway (Figure 1.1 B) are transcribed from DNA into mRNA, and then translated by a ribosome into a polypeptide which carries an N-terminal signal peptide sequence. This signal peptide consists of a tripartite structure: a) a positively charged N-terminal region of variable length, b) a 15-20 amino acid hydrophobic region, and c) a polar C-terminal region containing a peptidase motif for cleavage of the signal peptide upon export (16). Within the N-terminal region of the signal peptide, the Tat-consensus motif is defined by the amino acid sequence S-R-R-x-F-L-K, where the characteristic and invariant twin-arginine residues precede a variant amino acid (x) and the hydrophobic region.

Cytosolic chaperones, such as DmsD and TorD (47-51) will bind to the signal peptide, prevent targeting to the Sec pathway, protect the protein from premature signal peptide cleavage by peptidases, and assist the protein during its folding in the cytoplasm. Proteins targeted to the Tat pathway typically fold rapidly, attain their tertiary structure, associate and bind protein subunits (52,53) and/or redox cofactors

(54), such as FeS clusters or molybdopterin centers resulting in a Sec-incompatible protein within the cytoplasm (55).

After the substrate has matured (by binding to a chaperone and/or a cofactor), it is targeted to the TatBC complex at which point the signal peptide is released by the chaperone and "handed" over to TatC (39). Once the substrate-TatBC complex is formed, TatA is recruited, and through a yet unconfirmed step, the substrate is transported into the periplasm. Within the periplasm, the signal peptide is now exposed, revealing a signal peptidase-I recognition sequence which can be proteolytically cleaved (56) yielding the mature protein.

To understand the *E. coli* Tat pathway and the protein-protein interactions which occur within it, numerous experiments have been preformed both *in vivo* and *in vitro*, however not all results provide the clarity necessary to understand the dynamic interactions within the pathway. Therefore, we explored if bimolecular fluorescence complementation (BiFC) could be adapted for visualizing protein-protein interactions *in vivo* throughout the entirety of the *E. coli* Tat pathway.

1.3 Yellow Fluorescent Protein - Bimolecular Fluorescence Complementation (YFP-BiFC)

Many Tat pathway protein-protein interactions described previously were identified using traditional protein purification and identification methods (e.g. column chromatography (57), bacterial two-hybrid systems (1)). These techniques are either done *in vitro* or depend on a secondary reporter (e.g. fluorescent, enzymatic) to relay a signal that the protein interactions under interrogation occurred. YFP-BiFC eliminates the limitations of these techniques by detecting the protein interactions *in vivo* as well as providing a direct measure of the interaction through the emission of a fluorescent signal.

Numerous fluorescence-based methods have been developed for visualizing and identifying interacting proteins including FRET (58,59) and BiFC (60,61). In the case of YFP-BiFC (Figure 1.2 A), yellow fluorescent protein is split at a loop between two beta sheets, generating two non-fluorescent fragments, Y1 (aa 1-157) and Y2 (aa 157-238). When these two fragments are fused to a pair of interacting proteins (Figure 1.2 C), the native affinity between the proteins brings the split fragments into close proximity, resulting in the reassembly of the fluorescent protein. In this manner, the reconstituted fluorescence is directly coupled to the interaction between the two proteins and can be used to determine when and where *in vivo* the two proteins interact. Nyfeler *et. al.* 2005 (62) first demonstrated the YFP-BiFC technique in mammalian cells and we have applied this technique to monitor various protein interactions throughout the *E. coli* Tat pathway using a two-plasmid approach (Figure 1.2 B).

The YFP fragments, Y1 and Y2 were cloned into two plasmid vectors named p_POI_-Y1 and p_POI_-Y2, where proteins of interest (POI) could be cloned N-terminally in frame to the respective fragments. Coexpression of these plasmids in cells is possible due to their antibiotic resistance markers (ampicillin and kanamycin resistance, respectively), while plasmid copy number levels are regulated based on origins of replication (pBR322 and p15A, respectively), and protein expression is induced with the addition of a isopropyl β-D-1-thiogalactopyranoside (IPTG) in a titratable manner. Additionally a flexible linker, consisting of amino acids RSIAT or KQKVMNH for Y1 or Y2, respectively (60) was inserted between the POI and the YFP fragments to allow for a degree of flexibility between the fusion proteins. In all the experiments described hereafter, only the POI was altered; therefore changes in fluorescent signal intensity detected by YFP-BiFC are solely dependent on the two interacting proteins.

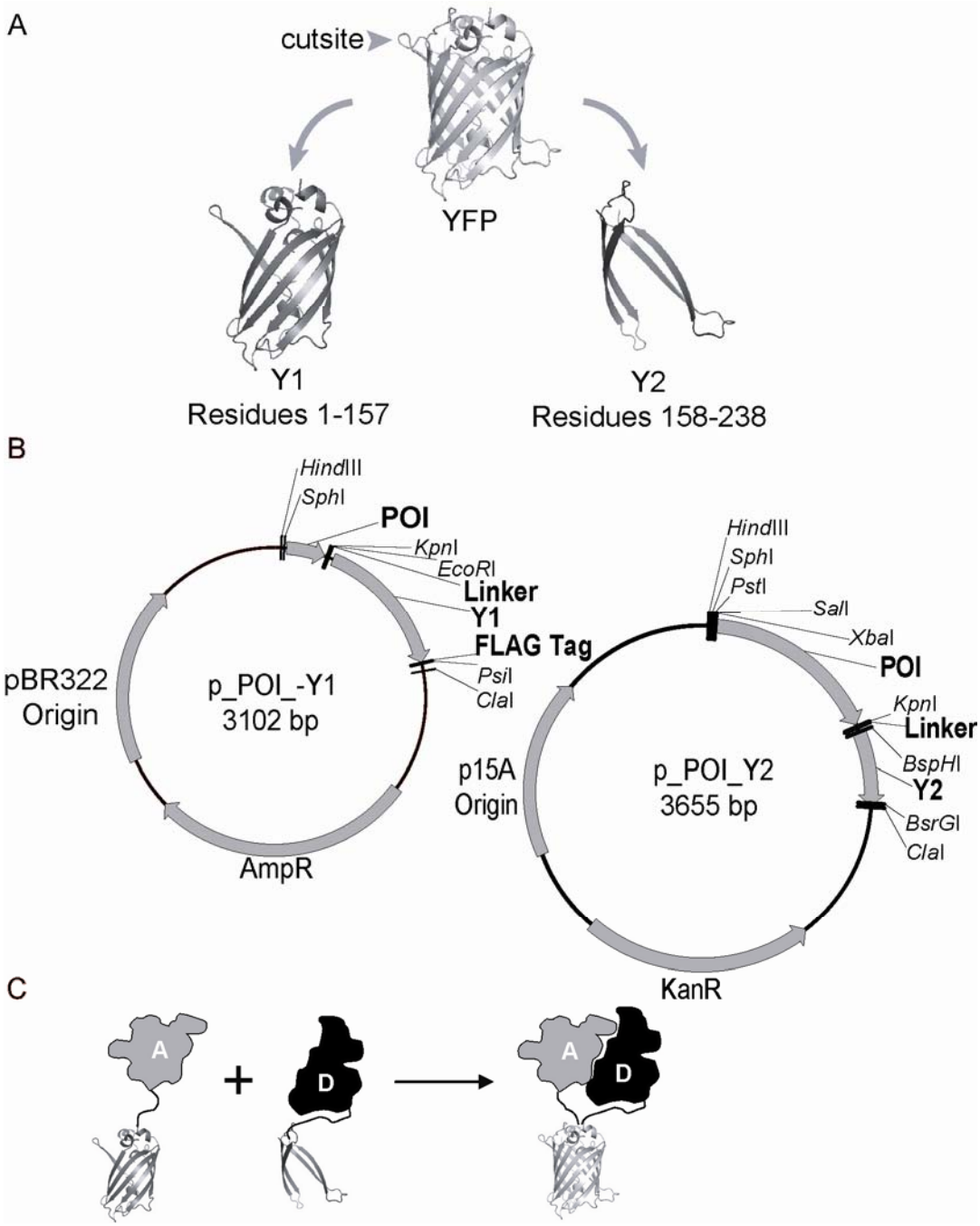

Figure 1.2 YFP-BiFC. (A) YFP is split between amino acids 157 and 158 to generate two fragments, Y1 and Y2, which are used for YFP-BiFC (PDB ID 1yfp). (B) The two plasmid system used to co-express fusions between proteins of interest (POI) and the Y1 or Y2 fragments. (C) Coexpression of protein A-Y1 with D-Y2 results in the refolding of the two YFP fragments, chromophore cyclization and fluorescence of YFP.

The benefits of using YFP-BiFC over other protein fragment complementation

assays (PCA) stems from: a) fluorescence is easily detected, it can be quantified, and fluorescent proteins have been well characterized by the scientific community; b) no cell permeable substrates (e.g. ampicillin for beta-lactamase) for enzymatic reporters need to be supplemented to the system to generate a detectable signal and c) fusion proteins will continue to interact with other cellular proteins without altering their native interactions (61,63-65).

However, there are caveats against using YFP-BiFC as a sole reporter for the detection of protein interactions. The first limitation of the assay lies in the irreversible association between the fluorescent fragments (60,62,66,67), which can inhibit native interactions between the proteins of interest if disassociation of one of the proteins is required for further downstream reactions. Since the YFP-BiFC assay is dependent on protein folding and subsequent fluorophore ring cyclization, the protein-protein interactions cannot be studied at real time rates (e.g. on the order of seconds or minutes), rather an hour needs to elapse before a detectable signal arises (68). Another limitation of this assay is the ability of the Y1 fragment to cause the fusion protein to destabilize and subsequently cause the protein chimera to be proteolytically degraded or is packaged into insoluble inclusion bodies (66,67). To overcome this destabilization effect of the Y1 fragment, fusion proteins to Y1 are expressed on a high copy vector allowing for a sufficient amount of protein to be expressed to allow for the proteins of interest to associate and subsequently generate a YFP-BiFC signal (60,66,69). Another limitation to the YFP-BiFC assay is the strength of affinity between the two interacting proteins. The affinity limit in terms of the disassociation constant (K_D) to generate a detectable YFP-BiFC signal was ~1mM when an interacting leucine zipper library was screened (70). The final limitation of this assay is the steric hindrances imposed by proteins to which the fragments are fused which prevents the fragments from interacting and subsequently folding to

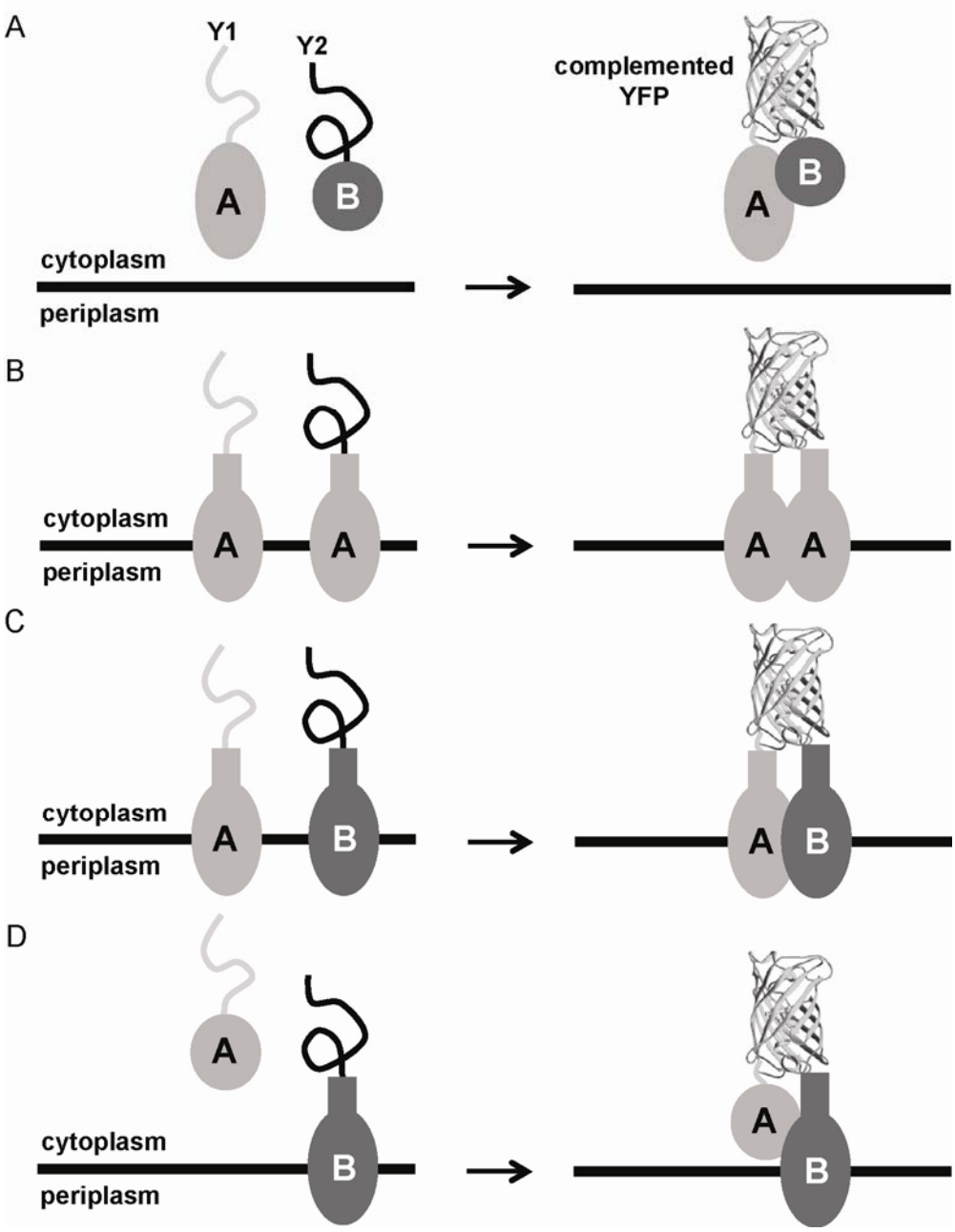

Figure 1.3 Types of protein-protein interactions detected via YFP-BiFC in the *E. coli* Tat Pathway. (A) Two soluble protein interactions in the cytoplasm, (B) homomeric transmembrane protein interactions, (C) heteromeric transmembrane protein interactions, (D) soluble cytosolic protein interacting with a transmembrane protein.

generate the fully folded YFP protein. To overcome this limitation, the directed design of linkers can be undertaken, but it is favorable to minimize external variables like various types of linkers and maintain a constant linker length between different

protein chimeras. Additionally extended linker lengths can promote proteolytic degradation and loss of the YFP fragments from their fusion partners.

The protein interactions that we wished to visualize with YFP-BiFC (Figure 1.3) *in vivo* in the *E. coli* Tat pathway are: (A) two heterologous soluble cytoplasmic proteins, (B) two homologous inner membrane anchored proteins, (C) two heterologous inner membrane proteins, and (D) a soluble cytoplasmic protein interacting with an inner membrane anchored protein. To demonstrate the applicability of YFP-BiFC in the *E. coli* Tat pathway, the first set of protein-protein interactions targeted was the interaction between two heterologous soluble cytoplasmic proteins, *E. coli* DmsA and its cognate binding chaperone DmsD.

1.4 Detection of substrate-chaperone interactions

The *E. coli* protein DmsA (a dimethyl sulfoxide reductase) and its cognate chaperone DmsD (49,57,71,72), are well characterized native Tat pathway proteins. The DmsD chaperone recognizes the DmsA twin-arginine signal peptide (57) and assists in the biogenesis and assembly of the DmsA enzyme (73). Binding of DmsD to the signal peptide provides specific advantages to DmsA: a) protection from premature cleavage of the Tat-signal peptide (50), b) DmsD assists in the maturation of DmsA by allowing molybdenum cofactor insertion, and c) the interaction serves as a proofreading step that prevents premature export of incompletely folded DmsA (74,75).

1.4.1 DmsA and DmsD interactions monitored with YFP-BiFC

Since the DmsA signal peptide (ssDmsA) alone is sufficient to interact with DmsD (57), we created protein chimeras whereby ssDmsA was fused to the Y1 fragment (ssDmsA-Y1), while DmsD was fused to the Y2 fragment (DmsD-Y2). Upon coexpression of the two constructs in wt TG1 *E. coli* cells, a strong YFP-BiFC signal was detectable with fluorescence microscopy (Figure 1.4 A, left panel) and with

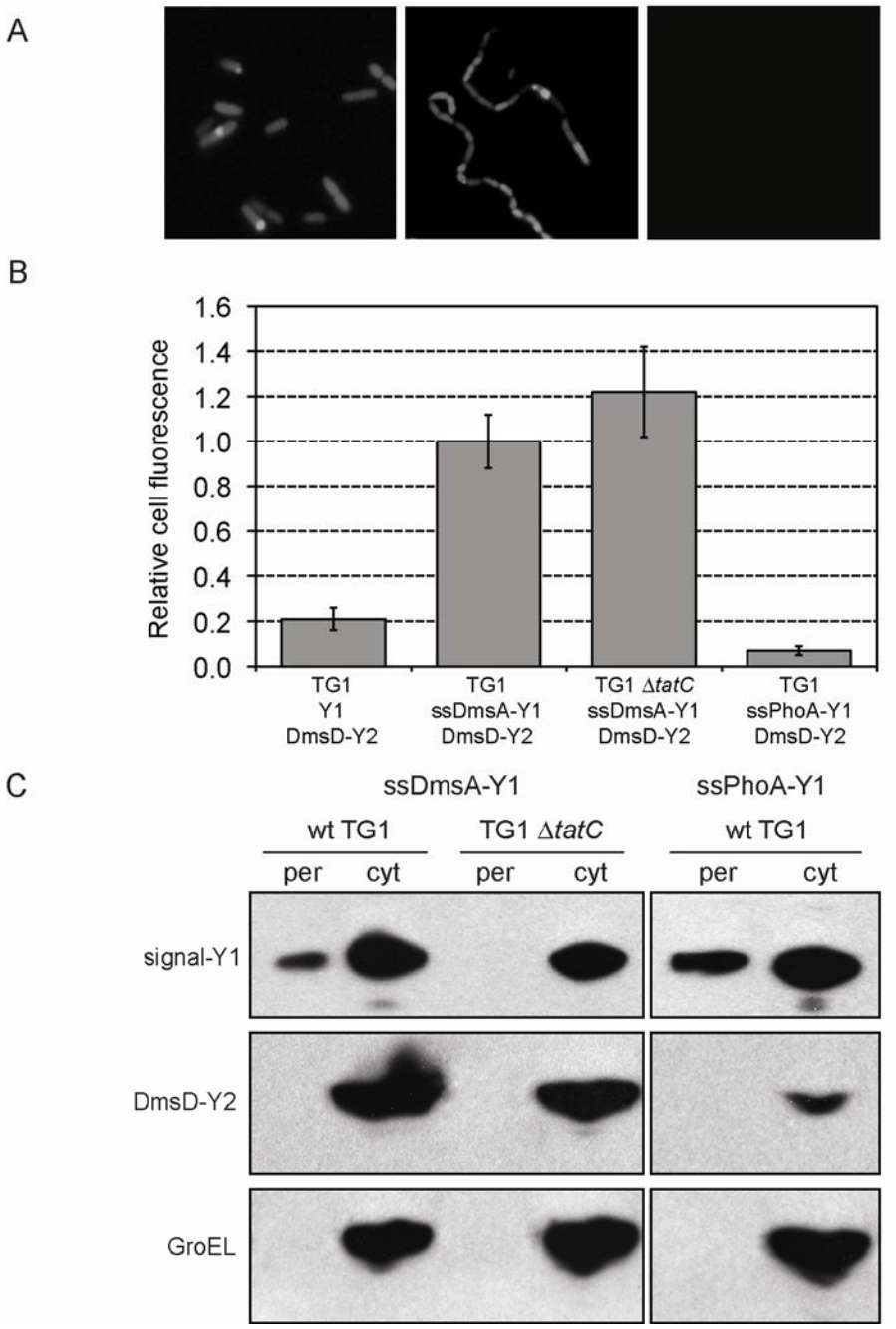

Figure 1.4 Proof of concept of YFP-BiFC between ssDmsA and DmsD. (A) Fluorescence microscopy images of ssDmsA-Y1/DmsD-Y2 coexpressed in either wt TG1 or TG1 Δ*tatC* cells (left panel, middle panel, respectively), and TG1 wt cells expressing ssPhoA-Y1/DmsD-Y2 (right panel) as a negative control. (B) FACS analysis of cells expressing indicated plasmid combinations from the average of 6 replicate experiments (n = 6). (C) Western blot analysis of periplasmic (per) or cytoplasmic (cyt) fractions from wt TG1 or TG1 Δ*tatC* expressing ssDmsA-Y1/DmsD-Y2 or ssPhoA-Y1/DmsD-Y2. Y1 was detected by antibodies against a C-terminal FLAG epitope tag, Y2 was detected using an anti-GFP, and GroEL was used as a fractionation marker and was detected using anti-GroEL.

fluorescence activated cell sorting (FACS) (Figure 1.4B) where the two interacting proteins generated a fluorescent signal 5 times brighter than the background control cells which were expressing an unfused version of Y1 and DmsD-Y2. The low levels of background fluorescence observed were likely due to self-assembly of the YFP fragments in the cytoplasm. When both ssDmsA-Y1/DmsD-Y2 were coexpressed in a Tat-pathway transport defective strain, TG1 Δ*tatC*, and equally strong fluorescent signal was observed (Figure 1.4 A and B) indicating the interaction between the two proteins was not dependent on a functional Tat-pathway. Importantly, when ssDmsA was replaced with a Sec-directed signal peptide, ssPhoA, and then coexpressed with DmsD-Y2 in wt TG1 cells, no fluorescent signal above the background negative control Y1/DmsD-Y2 (Figure 1.4 A and B) was detected. This result validates the YFP-BiFC signal observed was highly specific for the ssDmsA-Y1/DmsD-Y2 interaction and protein exclusivity between the Sec and Tat protein export pathways.

To determine if the chimeric protein ssDmsA-Y1 was still recognized and exported into the periplasm by the Tat translocase, we determined the subcellular localization of ssDmsA-Y1 when coexpressed with DmsD-Y2 in either wt or Δ*tatC* TG1 cells by Western blot analysis. As expected, ssDmsA-Y1 localized in the periplasm of wt TG1 cells, but was retained in the cytoplasmic fraction for Δ*tatC* TG1 cells (Figure 1.4 C – first column, top row). As for coexpression of ssPhoA-Y1 with DmsD-Y2, the protein chimera is detected in the periplasmic fraction, indicating export via Sec (Figure 1.4 C – second column, top row) and also in the cytoplasm demonstrating an overwhelmed Sec translocon. This result indicates that the lack of an observable YFP-BiFC signal for the coexpression of ssPhoA-Y1/DmsD-Y2 is not due to poor expression, or the highly efficient Sec translocon, rather it is due to lack of crosstalk between the Sec substrate and Tat chaperone.

1.4.2 DmsA and DmsD interactions monitored with RFP-BiFC

We next sought to determine if a non-engineered spectral variant of green fluorescent protein (GFP), specifically DsRED could be used for RFP-BiFC analysis within the *E. coli* Tat pathway. The rapidly maturing red fluorescent protein, DsRED, was initially isolated from the *Discosoma* sp. and through subsequent evolution a monomeric variant, mRFP1 Q66T, was isolated (76) with increased brightness and photostability. By examining the crystal structure of the protein (77) and comparing published RFP-BiFC split sites (76), we made two RFP fragments, termed R1 (aa 1-154) and R2 (aa 155-225) and replaced the Y1 and Y2 fragments in ssDmsA-Y1 and DmsD-Y2, respectively. Upon coexpression of ssDmsA-R1/DmsD-R2 in wt TG1 cells (Figure 1.5) we obtained nearly identical BiFC results as we had seen in the YFP-BiFC case. This result suggests that the signal detected via YFP-BiFC was due to the specificity and affinity between the substrate/chaperone interaction and not an artifact of the split reporter proteins.

1.4.3 DmsD-Y2-SsrA – a reporter for hitchhiker translocation

An open question was whether the ssDmsA-Y1 that was observed in the periplasm (Figure 1.4 C) was exported alone or was co-localized in a non-physiological complex with DmsD-Y2. This co-localization was a possibility because a) YFP-BiFC is an irreversible association, trapping the interacting proteins in a stable complex (66) and b) co-localization of partner proteins – termed hitchhiker translocation – has been observed for native and engineered Tat substrates (53,78). However, two lines of evidence demonstrate against co-localization of the substrate with the chaperone in the periplasm. The first piece of evidence, as seen in our fractionation data (Figure 1.4 C, second row), the DmsD-Y2 chimera was localized strictly in the cytoplasmic fraction. To further corroborate this fractionation data, we created a DmsD-Y2-SsrA chimera, where the SsrA tag targets the protein for

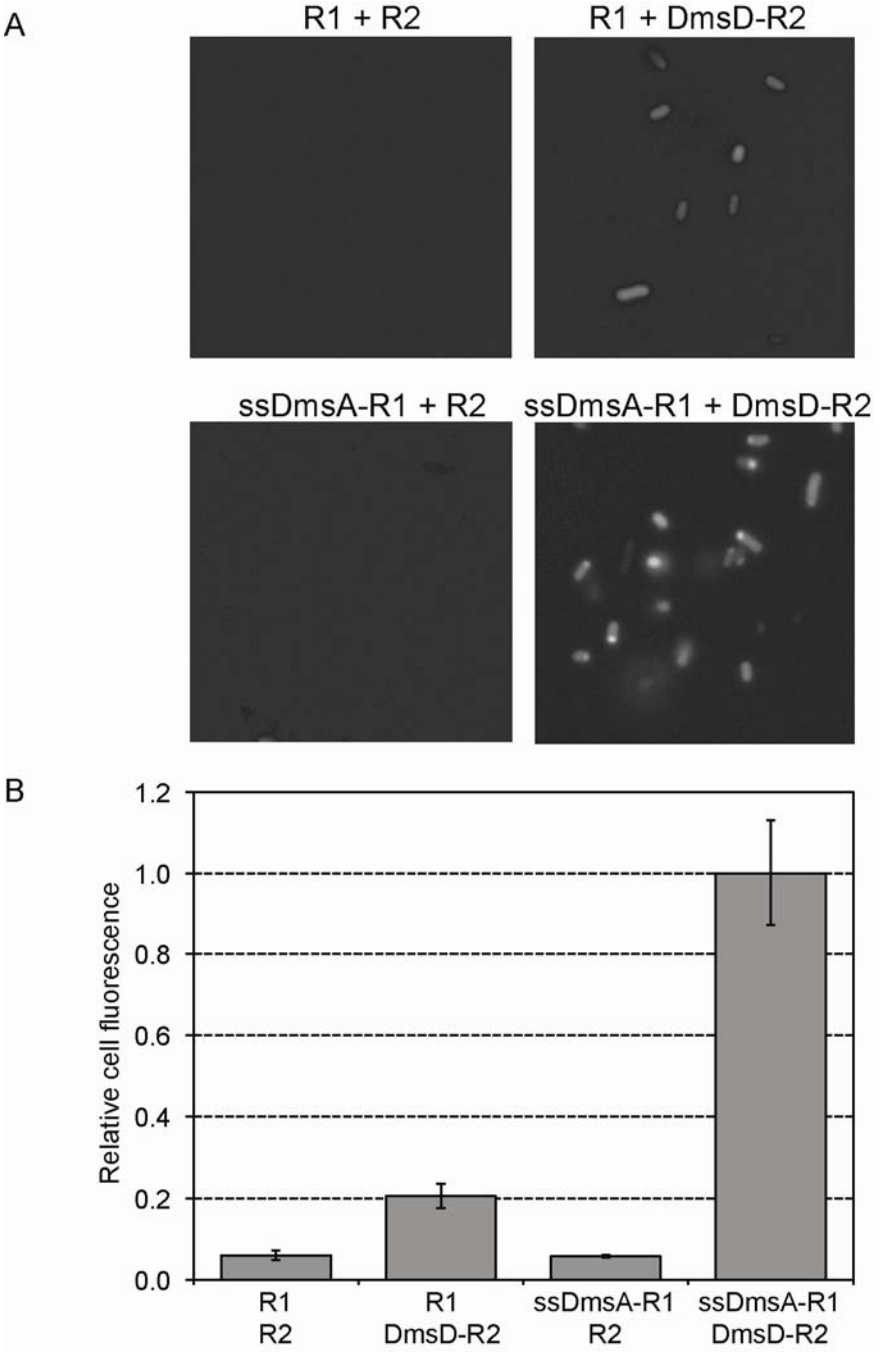

Figure 1.5 RFP-BiFC reports on ssDmsA DmsD interactions. (A) The fluorescent protein, mRFP1Q66T, was spilt into two fragments R1 (residues 1-154) and R2 (residues 155-238). Fluorescence microscopy images showing coexpression of ssDmsA-R1 with DmsD-R2 resulting in RFP BiFC (right bottom panel) with controls coexpressing the unfused R1 and/or R2 fragments as indicated. (B) Quantification of RFP BiFC signals using a microplate reader with fluorescence signal normalized to ssDmsA-R1/DmsD-R2 from the average of 3 replicate measurements (n = 3).

proteolytic degradation if present in the cytoplasm (79,80). If the coexpression of ssDmsA-Y1/DmsD-Y2-SsrA results in a Tat transport incompetent complex, a fluorescent signal would not be observed, however if the complex can be transported via Tat, a fluorescent signal should be observed that localizes to the periplasm.

Upon coexpression of either ssDmsA-Y1 or the full-length version of DmsA (DmsA-Y1) with DmsD-Y2-SsrA, no YFP-BiFC signal was observable under FACS analysis (Figure 1.6). As a positive control, the plasmid pssTorA-GFP-SsrA (81) demonstrated that a fluorescent signal can be obtained from the periplasmically localized GFP which was exported by the Tat pathway and avoided cytoplasmic degradation. The results from the Western blot and the DmsD-Y2-SsrA reporter protein demonstrate that hitchhiker translocation is disallowed when one of the subunits is a proofreading chaperone.

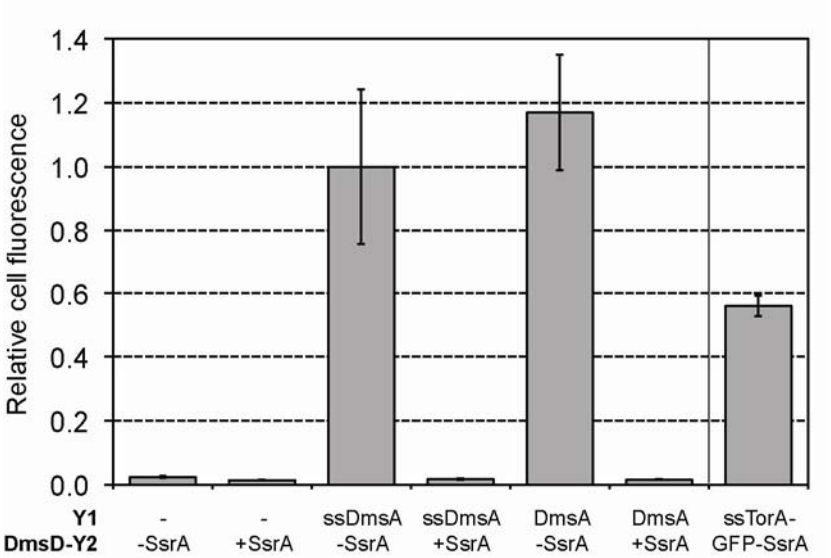

Figure 1.6 YFP-BiFC constructs localize in the cytoplasm. DmsD-Y2 with or without a C-terminal SsrA tag (+SsrA and –SsrA, respectively) was coexpressed with the negative control Y1(-), ssDmsA-Y1, or DmsA-Y1. Median fluorescence values obtained from FACS analysis with 3 replicate measurements (n = 3) and the data was normalized to the ssDmsA-Y1/DmsD-Y2 signal. The sstorA-GFP-SsrA plasmid was used as a positive control for the Tat-mediated rescue of SsrA-tagged GFP by export into the periplasm.

1.4.4 DmsA and DmsD YFP-BiFC occurs in the absence of TatABCE proteins

General cellular chaperones and specific Tat pathway proofreading chaperones typically bind to their substrates at an early stage of polypeptide synthesis. Specifically in the Tat pathway, this binding event is uncoupled from the membrane translocation step (75). As demonstrated previously (Figure 1.4), coexpression of ssDmsA-Y1 with DmsD-Y2 in a Δ*tatC* TG1 strain resulted in a fluorescent YFP-BiFC signal. We next wished to determine if the substrate-chaperone interactions required any specific part of the TatABCE proteins that make up the Tat-translocon or if this interaction was independent of these components.

Coexpression of ssDmsA-Y1/DmsD-Y2 in various *tat*-deficient TG1 strain backgrounds resulted in a significant YFP-BiFC signal increase even when the Tat pathway was partially (Δ*tatE*) or completely (Δ*tatAE*, Δ*tatB*, Δ*tatC*, and Δ*tatABCE*) inactivated in TG1 cells (Figure 1.7 A). We additionally constructed a full length DmsA chimera (DmsA-Y1) to determine if YFP-BiFC could be used to interrogate full length Tat substrates with their corresponding chaperone. Coexpression of DmsA-Y1/DmsD-Y2 resulted in an YFP-BiFC signal that was over background, but which was 50% the intensity of the ssDmsA-Y1/DmsD-Y2 interaction (Figure 1.7 B). The decrease in signal intensity is likely the result of steric hindrance caused by the larger DmsA-Y1 construct (110kDa versus 24kDa for ssDmsA-Y1, respectively) and the ability of DmsD to interact with other regions of full length DmsA (82).

It is noteworthy that coexpression of ssDmsA-Y1 or DmsA-Y1 with DmsD-Y2 in *tat*-deficient strains (Δ*tatAE*, Δ*tatB*, Δ*tatC*, and Δ*tatABCE*) resulted in a greater YFP-BiFC signal when compared to wt TG1 or Δ*tatE* TG1 cells. We believe this increase in fluorescence is due to the retention of ssDmsA-Y1 or DmsA-Y1 cytoplasmically, making all of the substrate available for YFP-BiFC complementation with DmsD-Y2, while in the wt TG1 or Δ*tatE* TG1 cells a portion of the substrate is exported into the

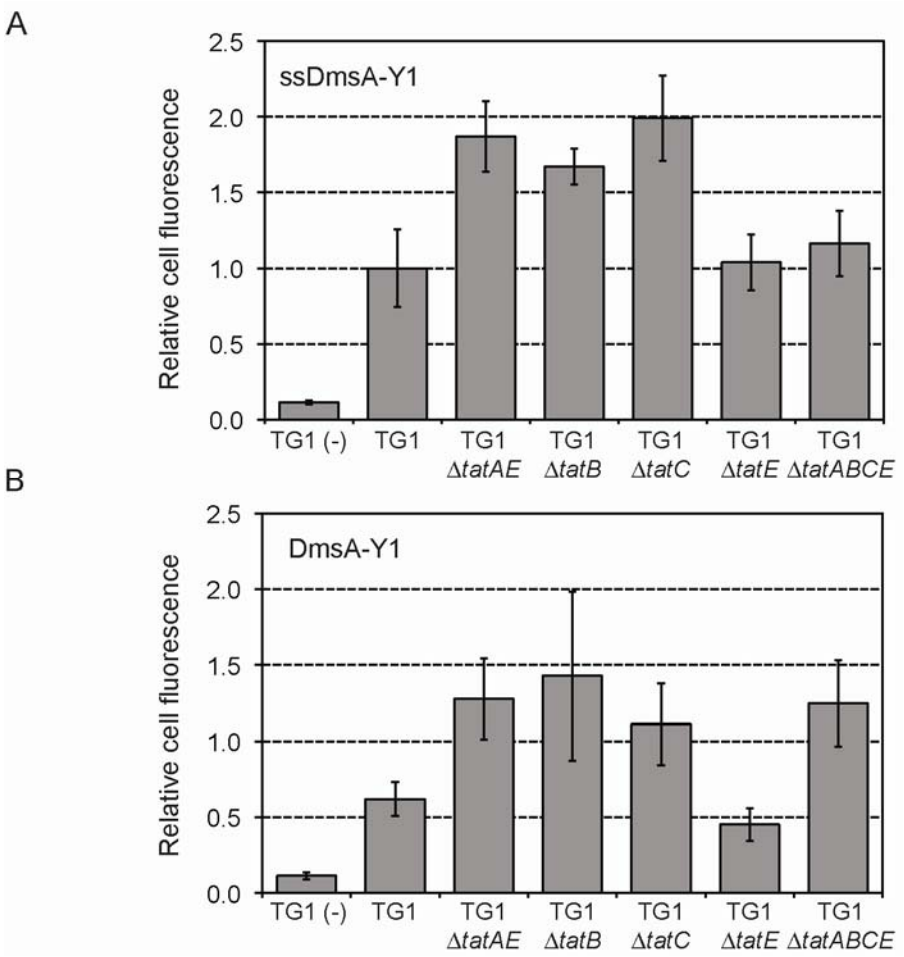

Figure 1.7 YFP-BiFC of DmsA and DmsD interacting in various Tat-deficient cells. (A) Coexpression of ssDmsA-Y1/DmsD-Y2 or (B) DmsA-Y1/DmsD-Y2 in various TG1 Tat deletion strains as indicated. Mean fluorescence was obtained via FACS and normalized to that for the wt TG1 ssDmsA-Y1/DmsD-Y2 interaction. Data is from the average of 6 replicate experiments (n = 6).

periplasm (Figure 1.4 C) since TatE is not an essential Tat pathway protein and TatA can substitute for the lack of TatE.

1.4.5 Twin-lysine mutations to DmsA

We next applied YFP-BiFC to determine if a functional Tat signal peptide was necessary to obtain a fluorescent signal. The twin-arginine residues in the Tat signal peptide consensus motif, S-R-R-x-F-L-K, were mutated to lysine residues (S-K-K-x-F-L-K), a substitution that completely abolishes export of Tat proteins (80,83). This same substitution, when applied to a truncated form of the ssTorA signal peptide *in*

vitro, had no effect on the binding affinity for the proofreading chaperone TorD when measured by isothermal titration calorimetry (ITC) (48).

We made the twin lysine mutations to both ssDmsA-Y1 and DmsA-Y1 generating ssDmsA(KK)-Y1 and DmsA(KK)-Y2 and coexpressed these constructs with DmsD-Y2 in wt TG1 cells. A strong YFP-BiFC signal was observed, however our *in vivo* analysis (Figure 1.8 A) demonstrated a 20-30% decrease in the fluorescence signal intensity between the twin-lysine when compared to the twin-arginine counterparts. This decrease in signal intensity between the twin-arginine and twin-lysine signal peptides can be attributed to the experimental conditions (*in vivo* versus *in vitro*) of this assay, to the utilization of the full length DmsA signal peptide or full length DmsA protein fused to Y1 and not a truncated variant of a Tat substrate (e.g. ITC experiments with ssTorA) and the affinity of the chaperone to the substrate which in turn affects the fluorescence intensity of the YFP-BiFC. Our results are consistent with the observations that the binding activity of proofreading chaperones is uncoupled from the membrane translocation step and the type of amino acid that is present in the canonical twin arginine motif of the signal peptide.

1.4.6 DnaK and crosstalk between Tat pathway chaperones and substrates

Other general molecular chaperones such as DnaK have been observed to interact with Tat substrates (57,84,85), presumably at an early stage of polypeptide synthesis and may participate in some form of protein proofreading. Coexpression of DnaK-Y2 with either ssDmsA-Y1 or DmsA-Y1 (Figure 1.8 A) resulted in a detectable YFP-BiFC signal in wt TG1 cells, however this fluorescent signal was much lower than that observed for the DmsD-Y2 interaction. Such a weak YFP-BiFC signal can be attributed to two potential factors, a) a less specific and/or a lower affinity of DnaK for DmsA compared to DmsD and b) tertiary structure differences (69kDa size of DnaK compared to 24kDa size of DmsD) that cause steric hindrances between the

YFP fragments inhibiting YFP-BiFC.

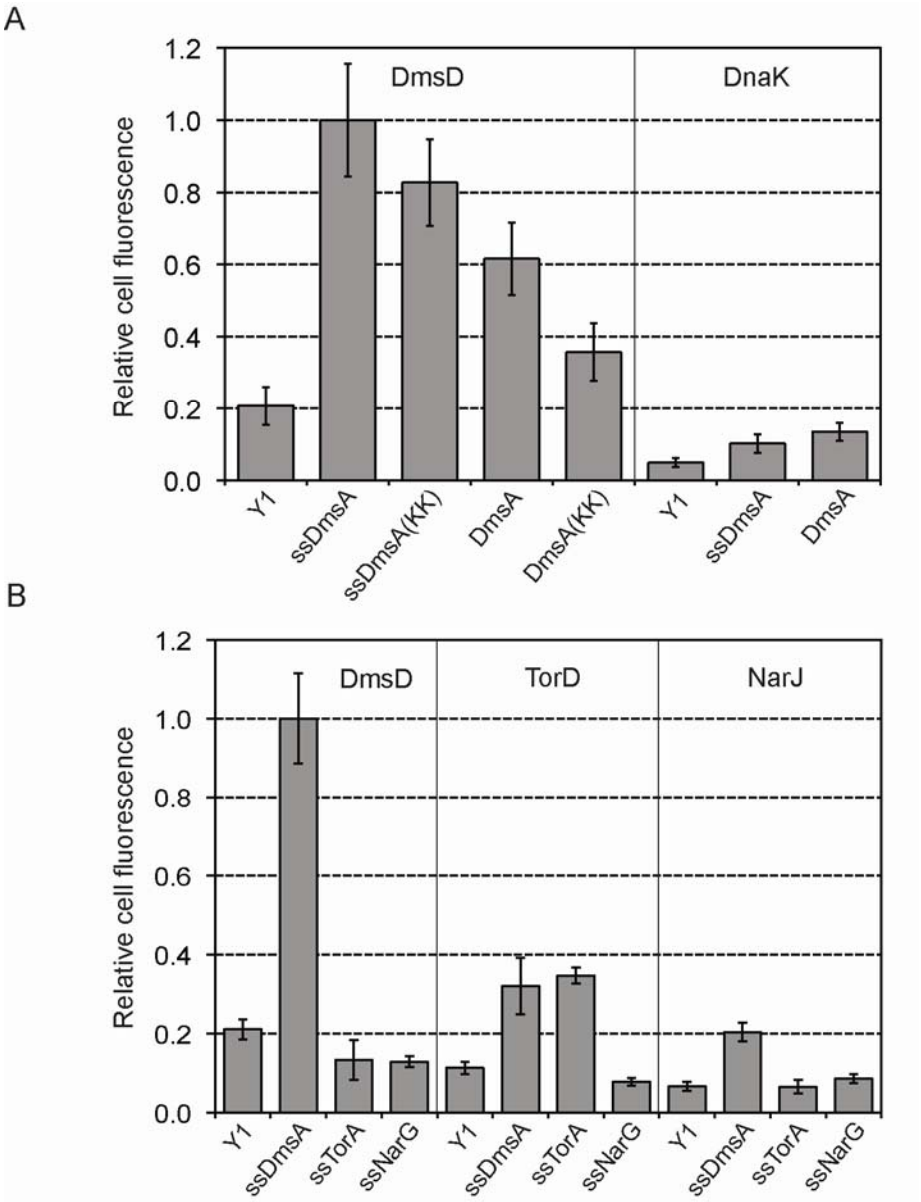

Figure 1.8 Interactions between various Tat pathway substrates and chaperones. (A) Coexpression of DmsD-Y2 with wt ssDmsA-Y1, full length DmsA or twin-lysine (KK) variants of ssDmsA-Y1 or DmsA-Y1 in a wt TG1 background. DnaK-Y2 was also co-expressed with ssDmsA-Y1 or DmsAY1. Y1 coexpressed with DmsD-Y2 or DnaK-Y2 served as negative controls. (B) Chaperones DmsD, TorD and NarJ fused to Y2 were coexpressed with their cognate or non-cognate signal sequences fused to Y1 (ssDmsA, ssTorA, ssNarG). Median cell fluorescence obtained via FACS a averaged from 6 replicate experiments (n = 6) and all data was normalized to the wt TG1 cells coexpressing ssDmsA-Y1/DmsD-Y2.

In addition to the aforementioned interacting protein pair DmsA/DmsD, additional substrate/chaperone pairs have been identified in the Tat pathway; these include TorA/TorD and NarG/NarJ. As with DmsD, TorD and NarJ have been experimentally shown to bind the signal peptide region of their respective substrates (TorA and NarG, respectively) (86-88). Although these chaperones have specificity for their own signal peptides, there have been documented reports of chaperone promiscuity when the cognate chaperone is missing for a substrate, as in the case where DmsD is able to bind both preprotein forms of DmsA and TorA (57). Utilizing the YFP-BiFC system developed for DmsA/DmsD interactions, we probed the specificity of the TorD and NarJ chaperones for their own cognate substrates and also for their ability to recognize non-cognate substrates (Figure 1.8 B).

The signal peptides of TorA and NarG were fused C-terminally to the Y1 fragment (creating ssTorA-Y1, ssNarG-Y1), while the chaperones, TorD and NarJ were fused C-terminally to the Y2 fragment (creating TorD-Y2, NarJ-Y2). Coexpression of all three chaperones with their cognate or non-cognate substrates in wt TG1 cells resulted in all three chaperones cross reacting with ssDmsA-Y1 to varying degrees with a YFP-BiFC fluorescent signal intensity on the order DmsD >> TorD > NarJ. The signal intensity of ssTorA-Y1/TorD-Y2 YFP-BiFC was only 35% of that for ssDmsA-Y1/DmsD-Y2 while no detectable fluorescent signal was observed above the negative control for the ssNarG-Y1/NarJ-Y2 YFP-BiFC. This lack of fluorescence for the interaction between ssNarG/NarJ, can be attributed to the differences in experimental methods and conditions (*in vivo* detection versus *in vitro*) and to the stability of the substrates (engineered chimeras versus native proteins) when they are fused to the YFP fragments.

Overall, this data corroborates previously published evidence that the Tat pathway chaperone TorD can cross associate with non-cognate signal peptides and

create a degree of protein redundancy within the Tat pathway. This data additionally demonstrates the flexibility and dynamic sensitivity of the YFP-BiFC assay where proteins of interest can be easily interrogated and their interacting partners can be evaluated in an *in vivo* setting.

1.5 Detection of machinery-machinery interactions

After thoroughly interrogating the soluble protein interactions that occur in the cytoplasm of the *E. coli* Tat pathway we now turn the YFP-BiFC assay to monitor the transmembrane protein components (TatA, TatB, and TatC) that make up the Tat translocon and which are essential for Tat export (see Figure 1.3 B and C). As mentioned previously, deletion of any component of the Tat translocon will result in a phenotypic change from single rod-shaped cells to elongated filamentous chains due to the cell wall amidase *lpxC* that is not exported into the periplasm by the Tat pathway (30). Since we previously made and tested YFP-BiFC in Tat machinery deletion strains (e.g. ssDmsA-Y1/DmsD-Y2 in Figure 1.7) our first set of experiments was to investigate if the chain phenotype could be complemented and single rod-shaped cells could be detected by brightfield microscopy analysis. We therefore tested the ability of each Tat component when fused to a YFP fragment (either Y1 or Y2) to complement the chain phenotype in single Tat machinery knockout strains (Figure 1.9) by expressing the chimeric proteins in *trans* from a plasmid.

Fusion of the Y1 fragment to any of the Tat machinery components (e.g. TatA-Y1) resulted in the inability to complement the chain phenotype. However, fusion of the Y2 fragment, to either TatA-Y2 or TatC-Y2, or coexpression of the two fragments on homologous machinery components (TatA-Y1/TatA-Y2, TatC-Y1/TatC-Y2) results in the restoration of the singlet-rod-like cell phenotype seen for wt TG1 cells. It is important to note, that neither TatB-Y2 nor coexpression of TatB-Y1/TatB-Y2

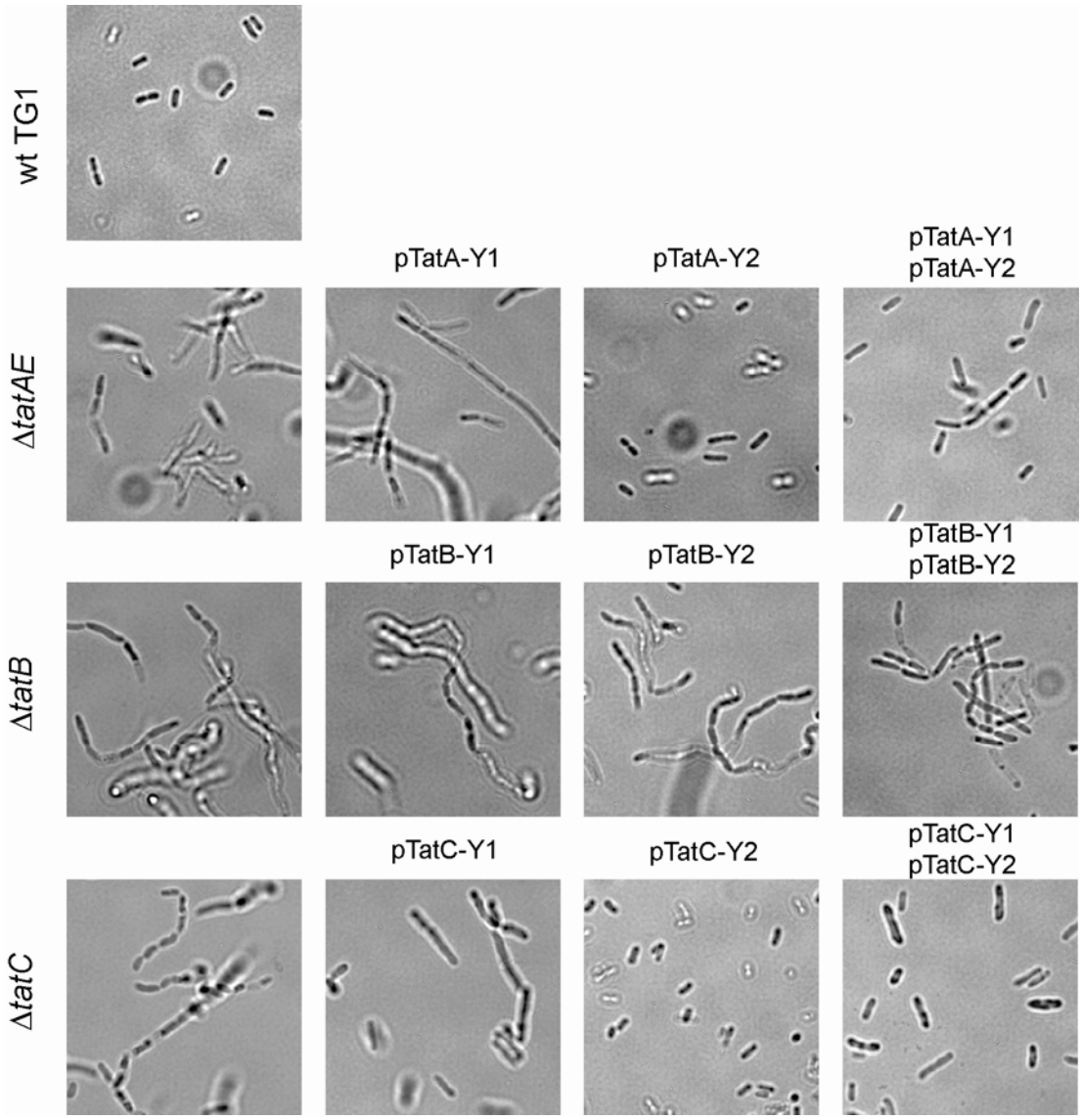

Figure 1.9 Phenotype complementation of Tat mutants when expressing Tat-machinery YFP fragment fusions. Brightfield microscopy images of wt TG1, TG1 *ΔtatAE*, TG1 *ΔtatB*, and TG1 *ΔtatC* cells (left panels). Expression of either Y1 or Y2 fragments fused C-terminally to TatA, TatB, or TatC in the respective Tat mutants (center two panels). Coexpression of both Y1 and Y2 fragments fused C-terminally to TatA, TatB, or TatC (right panel).

could restore the singlet-rod-like phenotype. This chain phenotype persistence in Δ*tatB* TG1 cells could be a result of gene dosage effects since it is well known that TatB and TatC associate at near equal levels; both at native or over-expressed levels, when analyzed by *in vitro* methodologies (40,89-91). With the Tat machinery

complementation data in hand, we applied YFP-BiFC to investigate the self-assembly of each Tat component in full knock out Δ*tatABCE* TG1 cells. Previous experimental evidence has demonstrated that the self-assembly of individual Tat components in strains lacking all other *tat* genes is not dependent on the presence of other Tat components (92-94), however to obtain a more dynamic understanding of the Tat pathway *in vivo* through YFP-BiFC, we included the single knock out strains (Δ*tatAE*, Δ*tatB*, ΔtatC0) in our experiments. Unless noted, the single knock out strains resulted in similar YFP-BiFC signals and results were reported for only the Δ*tatABCE* experiments. TatA was our first Tat component for YFP-BiFC analysis.

1.5.1 YFP-BiFC detection of TatA and F39A homo-oligomers

The *E. coli* TatA protein has been hypothesized as the channel through which substrates are exported into the periplasm and previous characterization studies have shown it to form stable, defined, homo-oligomeric complexes (41,42,95). A Δ*tatAE* TG1 strain was created to ensure that only the chimeric TatA supplemented to the cells was plasmid based since TatA and TatE share 50% amino acid homology and have been shown to have similar protein export functions (96).

As mentioned previously, we created TatA fusion chimeras to the YFP fragments (TatA-Y1 and TatA-Y2), and coexpressed these constructs in either Δ*tatAE* or Δ*tatABCE* TG1 cells. A fluorescent YFP-BiFC signal was detected, that localized to the polar regions (97) of Δ*tatAE* when observed under fluorescence microscopy (Figure 1.10) and when expressed in either Δ*tatAE* or Δ*tatABCE* had a signal intensity 2-3 fold above the negative controls (Figure 1.11). The YFP-BiFC signal intensities were roughly an order of magnitude lower than the observed signals for the ssDmsA-Y1/DmsD-Y2 interactions.

We also investigated a well characterized TatA mutant, F39A that has a single amino acid substitution in a predicted amphipatic region which blocks translocation

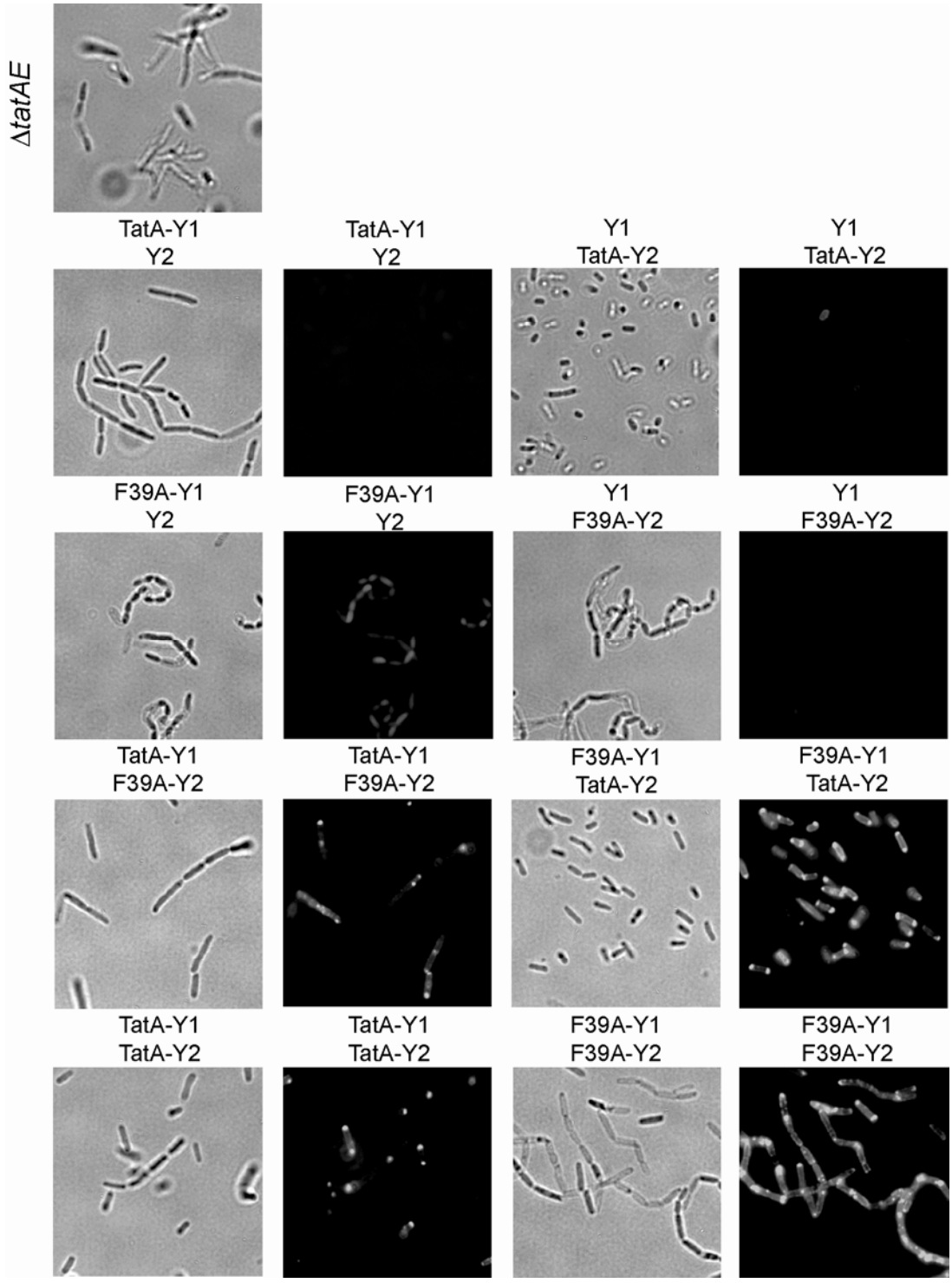

Figure 1.10 Visualizing the formation of TatA homo-oligomers with YFP-BiFC. Brightfield and fluorescence microscopy for phenotypic analysis of chain complementation and fluorescence localization in TG1 *ΔtatAE* cells expressing various TatA chimeras as indicated. Negative controls first two rows (e.g. TatA-Y1/Y2, F39A-Y1/Y2). F39A is *E. coli* TatA with a F39A mutation.

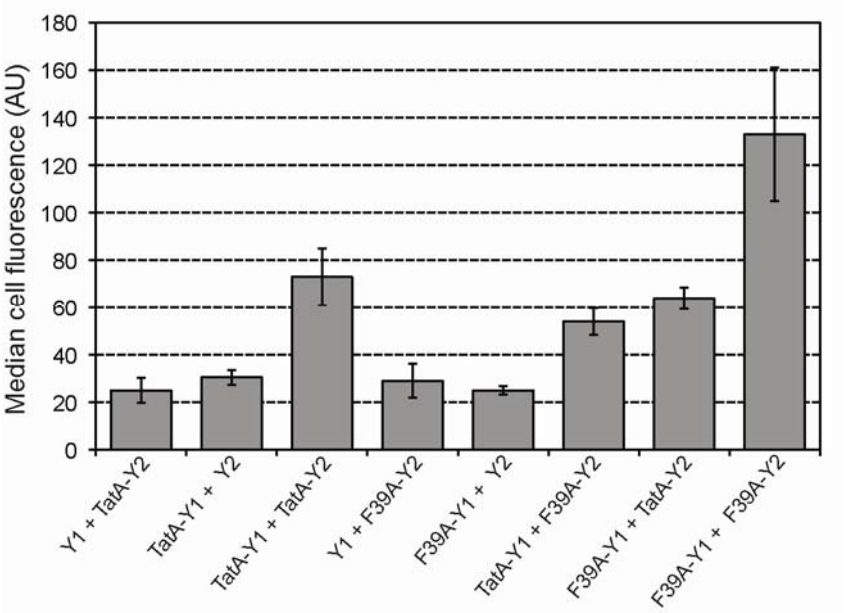

Figure 1.11 Quantitative measurement of YFP-BiFC fluorescence for Tat homo-oligomers. Cell fluorescence of TG1 Δ*tatABCE* cells expressing TatA-Y1, TatA-Y2, F39A-Y1, F39A-Y2 and the negative controls Y1 or Y2 in various combinations shown above. Median fluorescence values were obtained via FACS and reported as the average of 3 replicate measurements ($n = 3$).

and leads to aberrant TatA homo-oligomer formation (36). The F39A mutation was created in the TatA YFP-BiFC chimeras, generating F39A-Y1 and F39A-Y2, and when expressed in Δ*tatAE* cells could not complement the chain phenotype (Figure 1.10). Mixing the wt TatA-Y2 and F39A-Y1 mutant in Δ*tatAE* cells resulted in singlet cells with polar fluorescence localization, which was seen previously in the wt TatA YFP-BiFC.

Coexpression of F39A-Y1/F39A-Y2 in either Δ*tatAE* or Δ*tatABCE* TG1 cells resulted in a 2-fold increase in YFP-BiFC signal over the TatA-Y1/TatA-Y2 (Figure1.11), a brighter signal intensity was also seen under fluorescence microscopy (Figure 1.10), however these complexes were non-functional since the chain phenotype was still present in Δ*tatAE* TG1 cells. From this evidence, the export defect which characterizes F39A mutants is not a result of the inability to self assemble.

1.5.2 Heterologous Tat machinery interactions detected with YFP-BiFC

As in the YFP-BiFC analysis of TatA interactions, we created chimeras for TatB (TatB-Y1/TatB-Y2) and TatC (TatC-Y1/TatC-Y2) and coexpressed these constructions in Δ*tatABCE* TG1 cells. Only coexpression of TatB generated a YFP-BiFC signal (Figure 1.12); no fluorescent signal over the negative controls was detectable for TatC self assembly which is consistent with earlier observations where the overproduction of TatC in the absence of all other Tat components resulted in low levels of TatC incorporation into the inner membrane (94). The lack of YFP-BiFC signal for TatC self assembly is not due to the inactivity or instability of the chimeric proteins since the chain phenotype of Δ*tatC* is complemented resulting in singlet cells (Figure 1.9).

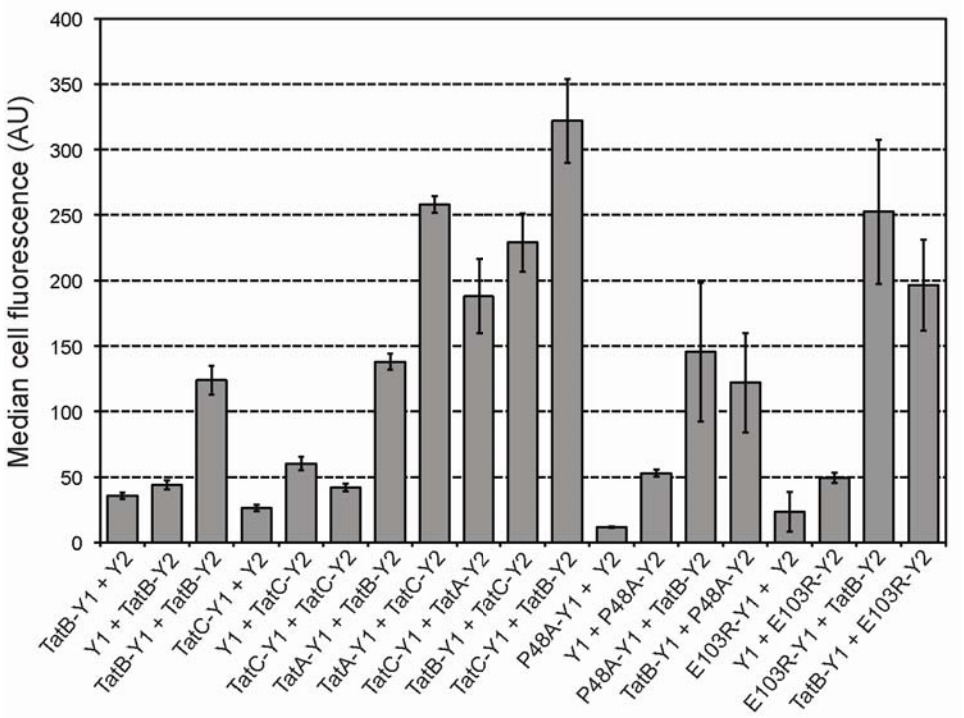

Figure 1.12 YFP-BiFC assembly of TatBC homo- and hetero-oligomers. (A) Cell fluorescence of TG1 Δ*tatABCE* cells coexpressing TatB and TatC chimeras as indicated above. In addition to wt TatC, TatC variants (P48A, E103R) were also evaluated for YFP BiFC. Unfused Y1 and Y2 were coexpressed with TatB, TatC, P48A, E103R and served as negative controls. Median fluorescence values were obtained via FACS and are the average of 3 replicate (n = 3) measurements.

Our next experiments were targeted at detecting heterologous Tat machinery interactions in Δ*tatABCE* TG1 cells. Coexpression of different pairs of Tat components (e.g. TatA-Y1/TatC-Y2) in Δ*tatABCE* TG1 cells resulted in measurable YFP-BiFC fluorescence signals for TatA-Y1/TatB-Y2, TatA-Y1/TatC-Y2 and TatB-Y1/TatC-Y2 (Figure 1.12) where the fluorescent signal obtained was 3-4 times that of the negative controls. The results obtained are consistent with previously published findings where TatB and TatC associate as equimolar subunit complexes (40) and that TatB can interact with TatA in the absence of TatC (42). Importantly, the YFP-BiFC signal can increase (TatC-Y2/TatB-Y1 versus TatC-Y1/TatB-Y2) or decrease (TatA-Y1/TatC-Y2 versus TatA-Y2/TatC-Y1) depending on which chimeric protein fusion is expressing the Y1 or Y2 fragments, indicating possible steric hindrances and/or some Tat machinery components are more amenable to a certain YFP fragment.

In addition to examining the wt TatC interactions using the YFP-BiFC chimera proteins, we tested two TatC variants, P48A and E103R. The P48A point mutation occurs in the first periplasmic loop region of TatC which abolishes substrate export and partially impairs TatC interaction with TatB (36), while the E103R point mutation lies in the first cytoplasmic loop between transmembrane helices II and III and blocks substrate export but does not affect the TatC interaction with TatB (98). The chain phenotype of Δ*tatC* TG1 cells when transformed individually or co-transformed with YFP-BiFC chimera constructs P48A (Y1 or Y2) or E103R (Y1 or Y2) could not be complemented and no detectable YFP-BiFC signal was observed either by FACS or fluorescence microscopy (Figure 1.14 C).

When we coexpressed the P48A variant with TatB (e.g. P48A-Y1/TatB-Y2) and monitored their interactions with YFP-BiFC in Δ*tatABCE* TG1 cells, a notably marked decrease in fluorescence was seen, confirming that the interaction between the two proteins was impaired (Figure 1.12). However, coexpression of the E103R

variant with TatB (e.g. E103R-Y2/TatB-Y2) resulted in near wt YFP-BiFC signals, again confirming previously published observations.

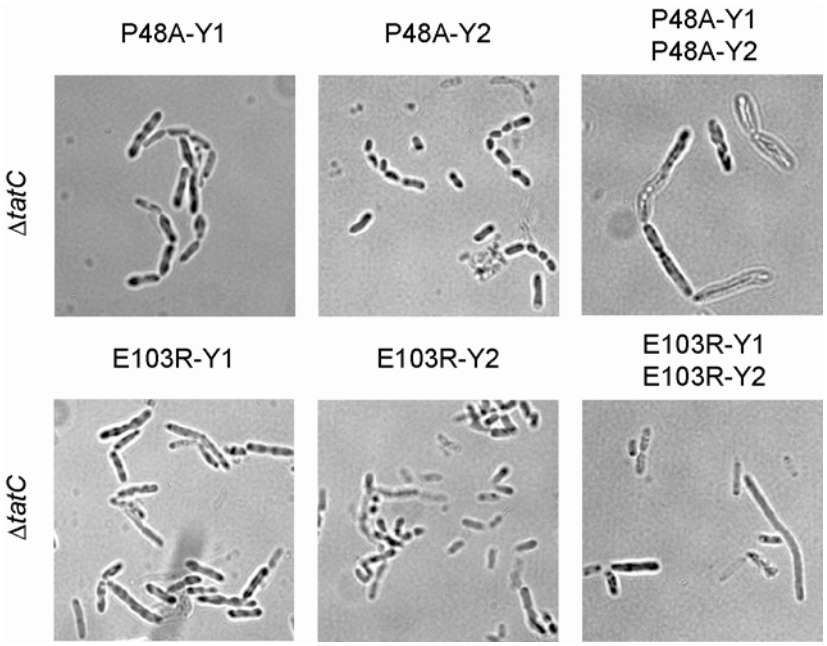

Figure 1.13 Expression of P48A and E103R YFP-BiFC chimera fusions in Δ*tatC* cells. Brightfield microscopy for phenotypic analysis of chain complementation in TG1 Δ*tatC* either for P48A or E103R TatC mutants fused to Y1 or Y2, expressed individual or together as indicated above.

1.5.3 TatBC YFP-BiFC interactions localize at cell poles

After obtaining a strong YFP-BiFC signal for TatB-Y1/TatC-Y2 and TatB-Y2/TatC-Y1 interactions in Δ*tatABCE* TG1 cells (Figure 1.12) we examined the cells using fluorescence microscopy and could see a bright fluorescent signal at the cell poles (Figure 1.14 A). We then co-transformed Δ*tatB* or Δ*tatC* TG1 cells with same plasmid combinations used for the Δ*tatABCE* experiment and we were able to see: a) the chain phenotype in both Δ*tatB* and Δ*tatC* TG1 cells was complemented resulting in singlet cells and b) the fluorescence signal from the YFP-BiFC was localized at the cellular poles (Figure 1.14 B and Figure 1.15 B).

The first result, complementation of the chain phenotype in ΔtatB or Δ*tatC* TG1 cells, supports previously seen experimental results with regards to the relative

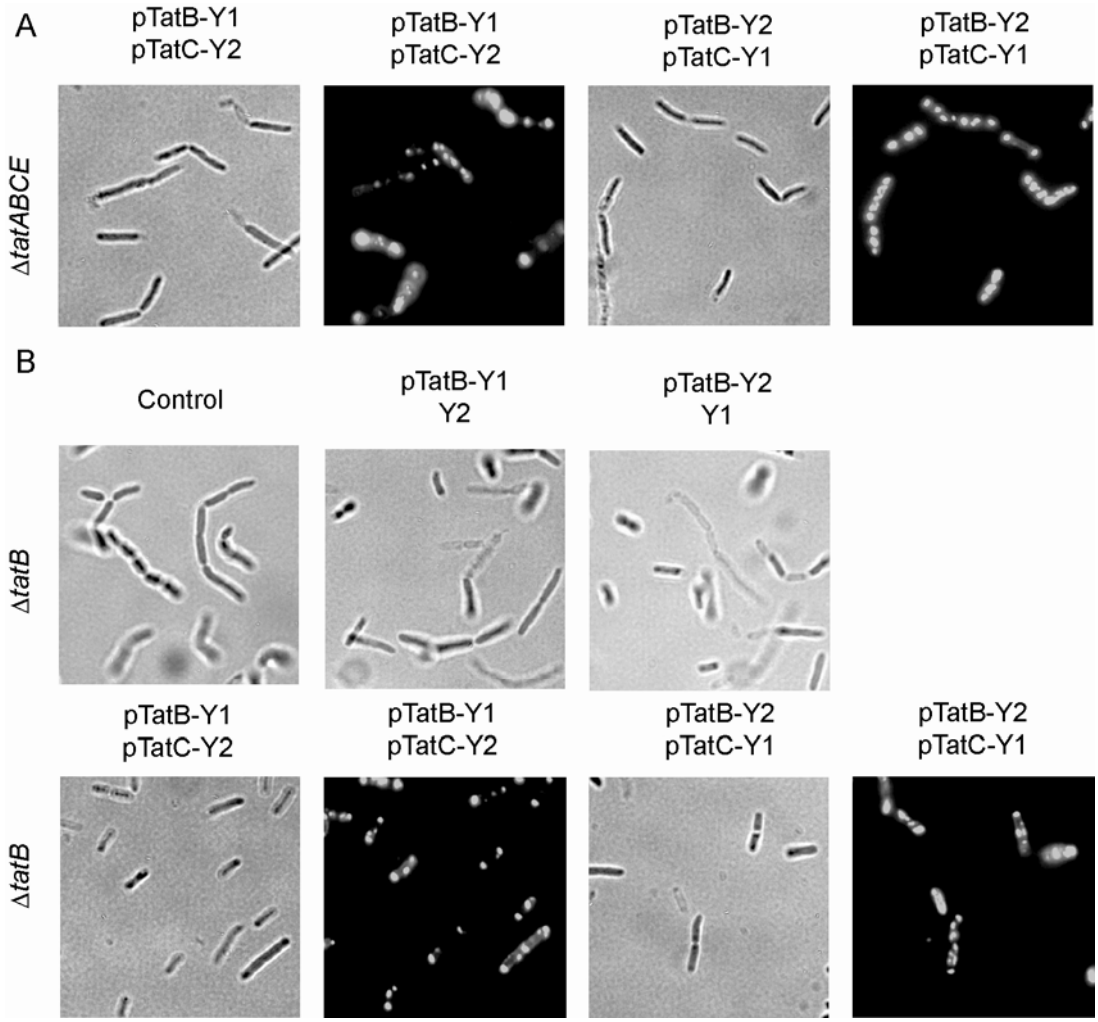

Figure 1.14 YFP-BiFC detection to TatBC hetero-oligomers in TG1 Δ*tatABCDE* and Δ*tatB* cells. (A) Brightfield and fluorescence microscopy of TG1 Δ*tatABCE* coexpressing either TatB-Y1/TatC-Y2 or TatB-Y2/TatC-Y1. (B) Brightfield microscopy for phenotypic analysis of TG1 Δ*tatB* as control cells or coexpressing TatB-Y1/Y2 or TatB-Y2/Y1 (middle row). Brightfield and fluorescence microscopy of TG1 Δ*tatB* cells for phenotypic analysis of chain complementation and fluorescence localization of coexpressed proteins TatB-Y1/TatC-Y2 or TatB-Y2/TatC-Y1 (bottom row).

amount of TatB and TatC, since overproduction of TatC alone will result in low levels of TatC integration into the inner membrane (94). Additionally, since the formation of a fluorescent YFP protein in the YFP-BiFC assay is dependent on a 1:1 ratio of TatB and TatC associating, this further supports published experimental evidence that these proteins exist at equimolar concentrations (40). Secondly, the localization of

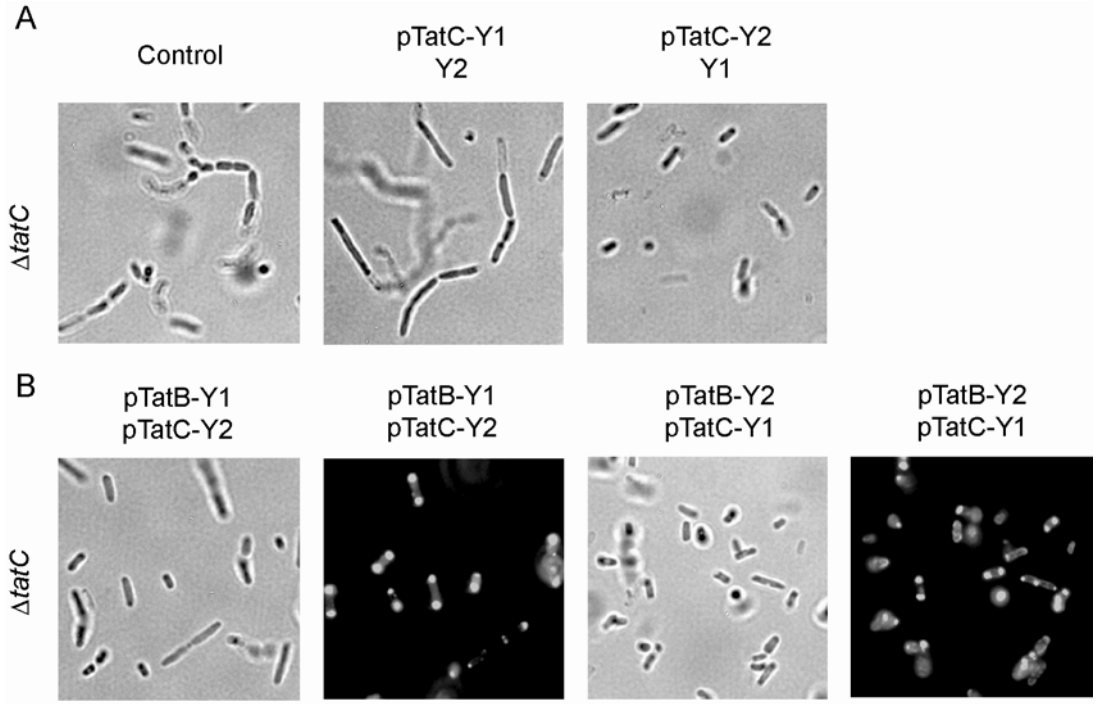

Figure 1.15 YFP-BiFC detection of TatBC hetero-oligomers in TG1 Δ*tatC* cells. (A) Brightfield microscopy for phenotypic analysis of TG1 Δ*tatC* as control cells or coexpressing TatC-Y1/Y2 or TatC-Y2/Y1. (B) Brightfield and fluorescence microscopy of TG1 Δ*tatC* cells for phenotypic analysis of chain complementation and fluorescence localization of coexpressed proteins TatB-Y1/TatC-Y2 or TatB-Y2/TatC-Y1.

fluorescence at the cellular poles when TatB and TatC chimeras are coexpressed for YFP-BiFC, gives the first *in vivo* experimental evidence that these proteins actually do associate at the poles. Previously published evidence was only able to identify the individual TatB and TatC proteins at the poles, but not the two in unison (97).

Through the directed application of YFP-BiFC to membrane bound proteins, the *in vivo* interactions between homologous and heterologous transport machinery components in the Tat pathway of *E. coli* could be observed. The dynamic range of the YFP-BiFC technique is sensitive enough to detect interactions between transmembrane proteins (e.g. TatA-Y1/TatA-Y2) or soluble cytoplasmic proteins (e.g. ssDmsA-Y1/DmsD-Y2), whereby the fluorescent signal intensity for the soluble interaction is an order of magnitude above that for the transmembrane interaction.

Additionally, we present the first experimental *in vivo* evidence that TatB and TatC co-localize at the poles of *E. coli*.

1.6 Detection of substrate-machinery interactions

After utilizing YFP-BiFC to detect soluble protein interactions in the cytoplasm (e.g. ssDmsA-Y1/DmsD-Y2) and Tat translocon proteins associated with the inner membrane (e.g. TatA-Y1/TatC-Y2), we sought to detect interactions between a soluble Tat substrate and the Tat translocon. The TatBC complex as examined with YFP-BiFC previously is known to interact with Tat pathway substrates (91,99) and act as a binding site before translocation into the periplasm. The ability of a Tat substrate to associate with the inner membrane and subsequently interact with the Tat machinery components has been shown to be independent of Tat associated chaperones [(100)], therefore our next set of YFP-BiFC experiments tested the ability of ssDmsA to interact with TatB or TatC components.

Using the previously designed YFP-BiFC protein chimera constructs, we coexpressed ssDmsA-Y1 with TatB-Y2, TatC-Y2, P48A-Y2 or E103R-Y2 and used FACS analysis to monitor the fluorescent signal from the interacting proteins (Figure 1.16) in a Δ*tatABCE* TG1 background. Fluorescence was observed between ssDmsA-Y1 and either TatB-Y2 or TatC-Y2, with an approximately 3.5 or 7 fold fluorescent signal above the negative control, respectively. Full length DmsA-Y1 was also coexpressed with either TatB-Y2 or TatC-Y2 and a much weaker but still significant YFP-BiFC signal was observed. No signal above background was observed when ssDmsA-Y1 was co-expressed with TatA-Y2.

To demonstrate a loss of YFP-BiFC, we took the two TatC variants described previously in Section 1.5.2, and we coexpressed ssDmsA-Y1 with either P48A-Y2 or E103R-Y2 in Δ*tatABCE* TG1 cells. The resulting YFP-BiFC signal was 1.5-2.0 fold

less fluorescent that that seen for the wt TatC interaction (Figure 1.16). These two mutations inhibit the export of substrates into the periplasm, which should cause a buildup of potential YFP-BiFC interactions and should generate a larger fluorescent signal. However this increase in fluorescence signal does not occur, and can be explained by recent experimental data

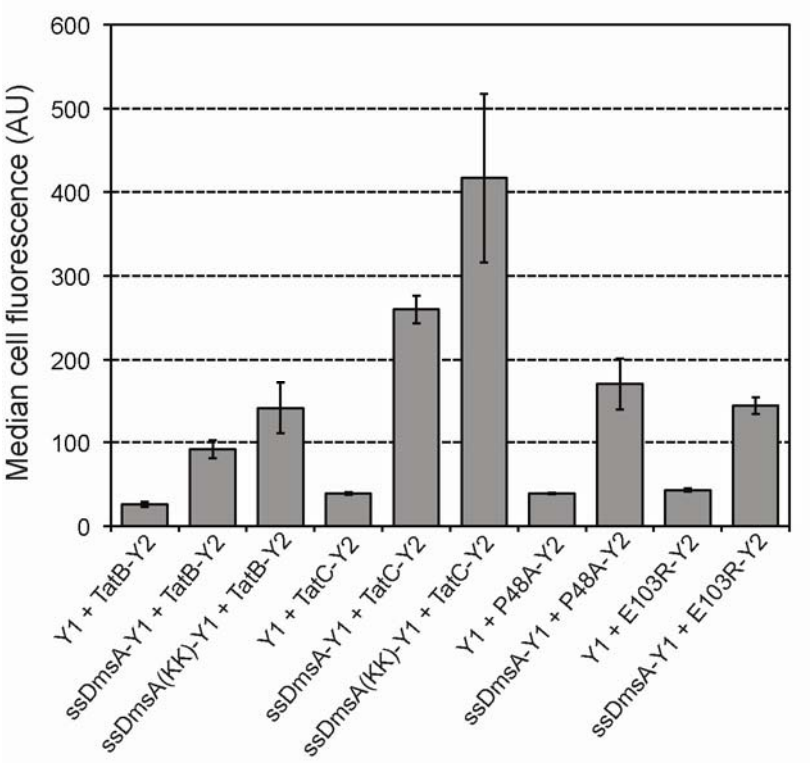

Figure 1.16 YFP-BiFC reveals ssDmsA interacting with TatB and TatC. (A) Cell fluorescence of TG1 Δ*tatABCE* cells coexpressing ssDmsA-Y1, ssDmsA(KK)-Y1 with TatB-Y2, TatC-Y2, P48A-Y2, or E103R-Y2 as indicated. Y1 unfused but coexpressed with Y2-chimeras was used as a negative control. Median fluorescence values were obtained via FACS and reported as the average of 3 replicate measurements ($n = 3$).

which indicates that E103R has a reduced binding affinity to substrates (98). We hypothesize this same reduced binding affinity exists for the P48A TatC variant resulting in a lower fluorescence signal. The decrease in fluorescent YFP-BiFC signal demonstrates that the YFP-BiFC is dependent on the affinity between the interacting proteins to generate a signal and not on the affinity between the YFP fragments.

Interestingly, when we tested the ability of ssDmsA(KK)-Y1 to interact with either TatB-Y2 or TatC-Y2 we observed a marked increase in YFP-BiFC signal that was 5.4- and 10.7-fold above background (Figure 1.16). This corroborates the recent observation that the twin-arginine residues of the Tat consensus motif are not essential for binding of the precursor substrate to the TatBC complex (98), however these residues are necessary for transport into the periplasm (101), demonstrating a highly specific motif recognition ability by the Tat pathway.

1.7 Detection of chaperone-machinery interactions

The last set of interactions that we wished to monitor with YFP-BiFC was between the soluble cytoplasmic proofreading chaperone DmsD and the TatBC components of the Tat translocon. Utilizing the previously constructed TatB and TatC YFP-BiFC chimeras, with one additional construct DmsD-Y1, we coexpressed these protein fusions in Δ*tatABCE* TG1 cells and detected a YFP-BiFC signal between DmsD, TatB and TatC (Figure 1.17).

As in the ssDmsA-Y1 YFP-BiFC experiment, the DmsD interaction with TatB or TatC resulted in a 2-4 fold increase in fluorescent signal above background. The YFP-BiFC signal, as in other cases tested, was shown to be negligibly affected when the orientation of the YFP fragments was changed (e.g. DmsD-Y1/TatC-Y2 versus DmsD-Y2/TatC-Y1. Interestingly, when the DmsD chimera was coexpressed with either P48A or E103R, a dramatic decrease in YFP-BiFC signal over the wt TatC interaction was seen. This result suggest that mutations to either of these residues in TatC, affects the manner in which DmsD interacts with the Tat translocon and on a larger scale how the transfer of the substrate (e.g. DmsA) can be impaired and translocation can be blocked.

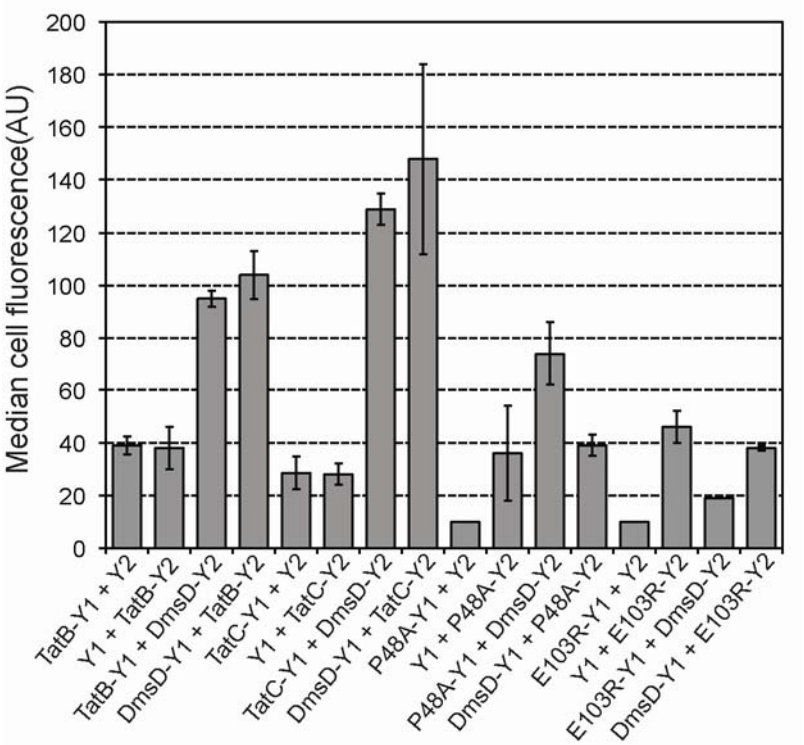

Figure 1.17 YFP-BiFC reveals DmsD interacting with TatB and TatC. Cell fluorescence of TG1 Δ*tatABCE* cells coexpressing DmsD-Y1 or DmsD-Y2, with TatB, TatC, P48A, or E103R chimeras as indicated. Y1 or Y2 unfused but coexpressed with the other protein chimeras was used as a negative control. Median fluorescence values were obtained via FACS and reported as the average of 3 replicate measurements ($n = 3$).

1.8 Discussion

Bacterial protein export requires a wide range of protein interactions between soluble and transmembrane proteins. The rate limiting step in understanding these protein pathways *in vivo* has been the development of a dynamic reporter that can probe these interactions directly. Here we have demonstrated that YFP-BiFC is such an *in vivo* reporter that has enabled us to detect numerous protein-protein interactions along the entirety of the Tat pathway in *E. coli*. Specifically, we have observed interactions between a) soluble cytoplasmic proteins, b) homologous or heterologous transmembrane proteins and c) soluble cytoplasmic proteins with transmembrane proteins.

Utilizing a simple two plasmid expression system for YFP-BiFC and varying the protein of interest, while keeping the linker region and YFP fragments constant, an unlimited number of interactions can be probed. The dynamic range of YFP-BiFC, in terms of fluorescence intensity, spans an order of magnitude (eg. ssDmsA-Y1/DmsD-Y2 versus ssDmsA-Y1/TatC-Y2) and could be expanded by a) altering the geometry of the flexible linker between the protein of interest and the YFP fragment and b) engineering the YFP fragments themselves for increased solubility and quantum yield.

With our proof of concept proteins, DmsA and DmsD, we validated that the application of YFP-BiFC to the Tat pathway is indeed feasible. We then expanded our experiments to validate previously determined interactions and also gained new insight into the ways that the *E coli* Tat pathway functions *in vivo*. Beginning with substrate and chaperone protein interactions, we were able to observe that the initial association is governed by signal peptide recognition, but there were measurable differences in the emitted fluorescence from YFP-BiFC for signal peptides versus full length protein interactions (e.g. ssDmsA versus DmsA). Additionally, substrate-chaperone associations are not limited to their cognate substrates, but there is a degree of flexibility and redundancy within the Tat pathway that allows for a cross talk to occur so that substrates can be directed towards the Tat translocon.

Within the Tat translocon, transmembrane protein interactions both homologous and heterologous were observed by YFP-BiFC. Association between the TatABC proteins was confirmed as previously reported, but more importantly the signal intensity from the YFP-BiFC varied depending on which two transmembrane proteins were being interrogated. Mutations to components of the Tat translocon demonstrated that the YFP-BiFC signal had a dynamic range, by increasing (e.g. F39A) or decreasing (e.g. P48A and E103R) in fluorescence intensity, respectively. Finally, for the first time experimentally, we were able to demonstrate *in vivo* through

YFP-BiFC that TatB and TatC co-localize to the poles of *E. coli*.

Our final set of experiments with YFP-BiFC probed the interactions between soluble cytoplasmic proteins and their association with transmembrane Tat translocon proteins. Both substrate (ssDmsA) and chaperone (DmsD) were able to interact with either TatB or TatC and upon visualization of the interaction by fluorescence microscopy, a fluorescent signal at the poles was observed. These findings collectively provide the first genetic evidence that DmsD may play a role as a targeting factor that delivers substrates to the TatBC complex. Additionally, when either DmsA or DmsD chimeras were coexpressed with the mutated TatC proteins (e.g. P48A and E103R), a decrease of YFP-BiFC fluorescence was observed, supporting the idea that TatC can interact with both substrate and chaperone before any translocation events occur. Delineating the temporal sequence of interactions between these proteins cannot be accomplished using the method of YFP-BiFC due to the irreversible association between the YFP fragments and the time required for the YFP chromophore to form.

In conclusion, we have validated previously characterized protein interactions and uncovered new evidence for protein associations and spatial localization in the *E. coli* Tat pathway using the *in vivo* method of YFP-BiFC.

1.9 Materials and Methods

1.9.1 Bacterial strains, growth and induction conditions.

The bacterial strains that were used in this study are described in Table 1.1. For cloning purposes *E. coli* MC4100 cells were grown aerobically in either liquid LB media or on solid LB media with agar added. For the BiFC assay, TG1 cells were made electrocompetent by standard methods (102), transformed with equal plasmid concentrations, and grown overnight on solid LB media and antibiotics (BD Diagnostic Systems) at 37°C.

The next morning individual colonies were picked from the plates, placed into 3-mL of liquid LB with antibiotics in 16-18mm culture tubes, and grown aerobically for 4 hrs at 37^{o}C and 200 rpm or until the OD_{600} ~0.5. Isopropyl β-D-thiogalactoside (IPTG) was added to a final concentration of 1mM for induction of protein expression, the culture was then moved to a room temperature (RT) incubator (20-24^{o}C) at 200 rpm for the next 8 hours. Fluorescence was only observed for cells grown at RT, not at 37^{o}C.

All single knock out TG1 Tat mutants were generated by P1 transduction from the Keio collection (103) and kanamycin resistance was removed as described (104). Strain TG1 Δ*tatABCE* was first created by P1 transduction of Δ*tatE*::Kan^{R} from the Keio collection; the kanamycin resistance was removed as described previously, and P1 transduction was done again from BW25113 Δ*tatABC::aac* (32) whereby the apramycin resistance was not be removed.

Antibiotic selection was maintained for all markers on plasmids at the following concentrations: ampicillin, 100μg/ml; chloramphenicol, 20μg/ml; kanamycin, 50 μg/ml; and tetracycline, 10 μg/ml.

Table 1.1 Strains used in this study.

Strain	Description	Reference
MC4100	F' *araD139* Δ(*argF-lac*)*U169 rpsL150* (Str^r) *relA1 flbB5301 deoC1 ptsF25 rbsR*	Laboratory stock
TG1	F' *traD36 lacI^q*Δ(*lacZ*) *M15 proA$^+$B$^+$ / supE* Δ(*hsdM-mcrB*)*5* (*rk - mk - McrB$^-$*) *thi* Δ(*lac-proAB*)	Laboratory stock
TG1 Δ*tatAE*	TG1 derivative lacking the *tatA* and *tatE* genes	This work
TG1 Δ*tatB*	TG1 derivative lacking the *tatB* gene	This work
TG1 Δ*tatC*	TG1 derivative lacking the *tatC* gene	This work
TG1 Δ*tatE*	TG1 derivative lacking the *tatE* gene	This work
TG1 Δ*tatABCE*	TG1 Δ*tatE* derivative with an apramycin marked deletion Δ*tatABC::aac*	This work

1.9.2 Construction of plasmids.

Proof of concept interacting proteins chosen for YFP-BiFC experiments are *dmsD* and its cognate substrate *dmsA*. *E. coli dmsD* was polymerase chain reaction (PCR) amplified and cloned into the *Xba*I and *Kpn*I sites of pKNT25 (105); generating pDmsDT25 that harbors a chimerical gene encoding *dmsD* fused to the T25 fragment of the catalytic domain of *Bordetella pertussis* adenylate cyclase. Similarly, plasmid pssDmsALT18 was constructed by cloning a PCR fragment encoding the signal peptide of *dmsA* (excluding the signal peptide cleavage site) into the *Pst*I and *Kpn*I sites of pUT18.

To establish the BiFC assay system, PCR fragments encoding the N-(1-157aa) and C-terminal (158-238aa) halves of the enhanced yellow fluorescent protein (YFP), which were abbreviated as Y1 and Y2 hereafter, were amplified from pIAF817YFP (a gift from Dr. Rolf Morosoli). Plasmids pDmsD-Y2 and pssDmsA-Y1 were constructed by replacing the T25 and T18 fragments in plasmids pDmsDT25 and pDmsALT18 with Y2 and Y1, respectively. The linker sequences used for the fusion proteins are designed on the basis of those used by Hu *et. al*. (106). Further plasmid

constructions used in this and subsequent studies are based on these two initial plasmids.

For the RFP BiFC assay, mRFP1 with the Q66T point mutation was split at the same amino acid location as described previously (107), generating two fragments abbreviated as R1 and R2. These fragments were cloned in place of the Y1 and Y2 fragments in the pssDmsA-Y1 and pDmsD-Y2 plasmids, respectively.

Typical PCR reaction volumes were 50μL and genes cloned into plasmid vectors were either amplified from laboratory stock plasmids or from *E. coli* genomic DNA. $Vent_R^{®}$ Polymerase (New England Biolabs Inc, MA) was used for PCR on a BioRad Gradient Cycler (BioRad, CA). DNA purification from agarose gel electrophoresis or enzymatic digestions was carried out using QIAquick Spin Kits (Qiagen, CA) and QIAprep Miniprep Kits (Qiagen, CA) were used for plasmid purification. Restriction digestions using required enzymes were carried out as directed in the literature (New England Biolabs Inc, MA) and 20μL ligation reactions were carried out using T4 DNA Ligase (New England Biolabs Inc, MA).

Electroporation of plasmid DNA was carried out using a Gene Pulser Xcell microbial system (BioRad, CA) following the preset bacterial transformation protocol (1.8kV, 25μF, 200Ω) for a 1mm-gap electroporation cuvette. Plasmid constructions were carried out in MC4100 cells and sequence verified by the Cornell Life Sciences Core Laboratories Center. All plasmids have been annotated and all plasmids used in this study are listed in Table 1.2.

Table 1.2 Plasmids used in this study.

Plasmid	Description	Reference
pUT18	Plasmid containing T18 fragment of the catalytic domain of *B. pertussis* adenylate cyclase; Amp^R	(105)
pKNT25	Plasmid containing T25 fragment of the catalytic domain of *B. pertussis* adenylate cyclase; Kan^R	(105)

Table 1.2 (Continued)

pssDmsALT18	*E. coli dmsA* signal peptide inserted into pUT18	This work
pDmsDT25	*E. coli dmsD* coding sequence inserted into pKNT25	This work
pssDmsA-Y1	pDmsALT18 with T18 sequence replaced by sequence encoding YFP N-terminal fragment; FLAG tag epitope at 3' end	This work
pDmsD-Y2	pDmsDT25 with T25 sequence replaced by sequence encoding YFP C-terminal fragment	This work
pY1	Control plasmid expressing Y1-FLAG; made by removing *dmsA* signal peptide sequence from pssDmsA-Y1	This work
pY2	Control plasmid expressing Y2; made by removing DmsD from pDmsD-Y2	This work
pssPhoA-Y1	pDmsA-Y1-FLAG with *dmsA* signal peptide sequence replaced by DNA encoding the signal peptide of the *E. coli phoA* gene	This work
pDmsA-Y1	pssDmsA-Y1 with *dmsA* signal peptide sequence replaced by full-length *E. coli dmsA*	This work
pssDmsA(KK)-Y1	pssDmsA-Y1 with RR to KK substitution	This work
pDmsA(KK)-Y1	pDmsA-Y1 with RR to KK substitution	This work
pDnaK-Y2	pDmsD-Y2 with *dmsD* replaced by *E. coli dnaK* sequence	This work
pTorD-Y2	pDmsD-Y2 with *dmsD* replaced by the *E. coli torD* sequence	This work
pNarJ-Y2	pDmsD-Y2 with *dmsD* replaced by the *E. coli narJ* sequence	This work
pssTorA-Y1	pssDmsA-Y1 with *dmsA* signal peptide sequence replaced by *E. coli torA* signal peptide sequence	This work
pssNarG-Y1	pssDmsA-Y1 with *dmsA* signal peptide sequence replaced by *E. coli narG* signal peptide sequence	This work
pTatA-Y1	pssDmsA-Y1 with full-length *E. coli tatA* sequence in place of *dmsA* signal peptide sequence	This work
pTatA-Y2	pDmsD-Y2 with the full length *E. coli tatA* sequence in place of *dmsD*	This work
pF39A-Y1	pTatA-Y1 with F39A substitution	This work
pF39A-Y2	pTatA-Y2 with F39A substitution	This work
pTatB-Y1	pssDmsA-Y1 with full-length *E. coli tatB* sequence in place of *dmsA* signal peptide sequence	This work

Table 1.2 (Continued)

pTatB-Y2	pDmsD-Y2 with the full length *E. coli tatB* sequence in place of *dmsD*	This work
pTatC-Y1	pssDmsA-Y1 with full-length *E. coli tatC* sequence in place of *dmsA* signal peptide sequence	This work
pTatC-Y2	pDmsD-Y2 with the full length *E. coli tatC* sequence in place of *dmsD*	This work
pP48A-Y1	pTatC-Y1 with P48A substitution	This work
pP48A-Y2	pTatC-Y2 with P48A substitution	This work
pE103R-Y1	pTatC-Y1 with E103R substitution	This work
pE103R-Y2	pTatC-Y2 with E103R substitution	This work
pDmsD-Y1	pssDmsA-Y1 with *dmsD* in place of *dmsA* signal peptide sequence	This work
pR1	pY1 with mRFP1 Q66T N-terminus in place of Y1; FLAG epitope at 3' end	This work
pR2	pY2 with mRFP1 Q66T C-terminus in place of Y2	This work
pssDmsA-R1	pssDmsA-Y1 with N-terminus of mRFP1 Q66T in place of *dmsA* signal peptide sequence	This work
pDmsD-R2	pDmsD-Y2 with C-terminus of mRFP1 Q66T in place of *dmsA* signal peptide sequence	This work
pDmsD-Y2-SsrA	pDmsD-Y2 re-cloned to express Y2 with a C-terminal SsrA tag	This work
pssTorA-GFP-SsrA	ssTorA-GFP-SsrA; Cm^R	(108)
pDmsD-Y2 Tet^R	Kan^R marker in pDmsD-Y2 replaced with Tet^R marker	This work
pY2 Tet^R	Kan^R marker in pY2 replaced with Tet^R marker	This work
pTatB-Y2 Tet^R	Kan^R marker in pTatB-Y2 replaced with Tet^R marker	This work
pTatC-Y2 Tet^R	Kan^R marker in pTatC-Y2 replaced with Tet^R marker	This work
pP48A-Y2 Tet^R	Kan^R marker in pP48A-Y2 replaced with Tet^R marker	This work
pE103R-Y2 Tet^R	Kan^R marker in pE103R-Y2 replaced with Tet^R marker	This work

1.9.3 DmsD-Y2-SsrA construction and expression.

Primer extension PCR was performed to attach the coding sequence for the

SsrA degradation tag (AANDENYALAA) to the 3' end of Y2 which was then cloned into the pDmsD-Y2 plasmid, replacing the original Y2 fragment, and creating the construct pDmsD-Y2-SsrA. pDmsD-Y2-SsrA was then co-transformed with Y1 carrying plasmids into TG1 wt cells and plated overnight as described previously. After induction as previously described, the cells were analyzed via FACS for fluorescence at 8 hrs, then harvested by centrifugation (10000 x g, 10min, 4ºC) and resuspended in the same volume of fresh LB with antibiotics and without IPTG. The cells were returned to the 200 rpm shaker at RT and allowed to equilibrate for another 8 hrs before a final FACS reading for fluorescence was taken.

1.9.4 Fluorescence measurements, microscopy, and image processing.

After induction of protein expression, FACS data was collected on a BD FACSCalibur System (Franklin Lakes, NJ) at 0 and 8 hours post induction (Figure 1.18). Samples for FACS readings were prepared by diluting 50 µl of live bacterial

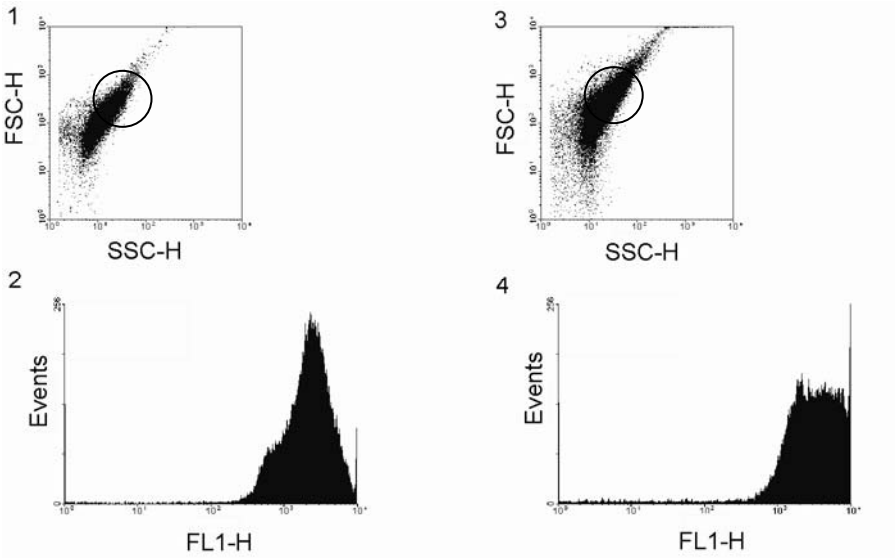

Figure 1.18 Gating of FACS data for TG1 wt and Δ*tatABCE* cells. Cell fluorescence collected by FACS for TG1 wt (panel 1 and 2) or TG1 Δ*tatABCE* (panel 3 and 4) cells coexpressing ssDmsA-Y1/DmsD-Y2. Panel 1 and 3 represent the forward and side scatter signals with the FACS gate (indicated by the black circle) that was used to count fluorescent cells. Panel 2 and 4 are histograms derived from panel 1 and 3, respectively, whereby these histograms are representative of the YFP-BiFC signal from the FL1-H FACS detector.

cells directly from culture in 1mL of 1x phosphate buffered saline (PBS). The FACS gate was then situated between the forward scatter detector (FSC-H) and side scatter detector (SSC-H) as indicated by the black circle for panels 1 and 3 in Figure 1.18. From the cells that passed through this gate, fluorescence measurements for YFP-BiFC were obtained by the FL1-H FACS detector. The fluorescence measurements were collected in replicates as indicated in the figures and the mean of the median fluorescence value was reported.

For RFP fluorescence measurements, the Biotek Synergy HT Plate Reader was used whereby 100µl of bacterial cells were spun down, washed twice with 1mL of 1x PBS, and resuspended in 150 µl of 1x PBS and placed into a Costar clear bottom-black side wall 96-well plate (Corning, NY). Fluorescence readings were taken with an excitation filter 530/25, emission filter 590/35, bottom detector setting, and sensitivity setting of 50.

For microscopy, 15µl of live bacterial cells directly from culture were placed onto a microscope slide with cover slip. All images were taken under oil immersion microscopy using a Zeiss 100x/1,30 lens. Microscopy was performed on a Zeiss Axioskop 40 equipped with a Zeiss 100x/1,30 Oil Plan-NEOFLUAR lens, a EXFO X-Cite light source (Mississauga, Ontario), a Semrock Brightline filter cube for YFP emission (YFP-2427A-ZHE) or RFP emission (Cy3-4040B-ZHE) (Rochester, NY), digitally imaged with a SPOT FLEX digital camera from Diagnostic Instruments, Inc. (Sterling Heights, MI) and controlled with Spot Imaging Software (Sterling Heights, MI). All images captured under 100x-oil immersion microscopy using the Zeiss 100x/1,30 Oil Plan-NEOFLUAR lens were under brightfield illumination (exposure 150ms) or under UV illumination (exposure 500ms). Images were saved as .tiff file format, imported into Adobe Photoshop (San Jose, CA), cropped and resized to 400x400 pixels, gray scale and auto leveled (bright field images only), imported into

Adobe Illustrator (San Jose, CA), scaled to 30% of 400x400 pixel size for the resulting figures.

1.9.5 Cellular fractionation and Western blotting.

After 8 hours of induction, 1ml of cells was taken and the OD_{600} was measured using a Biomate3 spectrophotometer (Thermo Scientific, MA). The cells were then spun down for 2 min at 13,000 x g and the supernatant was removed. The periplasmic fraction from the *E. coli* cells was isolated by using a modified protocol of the Epicentre Biotechnologies PeriPreps™ Periplasting Kit (Madison, WI), where the periplasting buffer did not contain any Ready-Lyse Lysozyme. The soluble protein fraction from the periplasmically treated *E. coli* cells was isolated with BugBuster® Master Mix (Novagen, Madison, WI) via the supplied Novagen protocol.

For Western blotting, an equal volume of 2x SDS-PAGE buffer was added to the periplasmic and soluble protein fractions and then boiled for 15 min at 100ºC. Samples were loaded onto 4-20% iGels (NuSep Ltd, Australia) where protein amount was normalized to the optical density of the cells taken before fractionation. After SDS-PAGE, the proteins were transferred to Immobilon-P PVDF 0.45µm membrane (Millipore, MA) and probed for the epitope FLAG (DYKDDDK) tag for Y1 fusions using the primary antibody Anti-FLAG® M2 (Stratagene, CA) with secondary antibody being Anti-Mouse IgG-HRP (Promega, WI). To detect the Y2 fragment, the primary antibody was anti-GFP (Roche, IN) with secondary Anti-Mouse as described above. As a cytoplasmic fractionation marker, the primary antibody Anti-GroEL (Sigma, MO) was used along with the secondary Anti-Mouse as described above. HRP detection was done by chemi-luminescence using the Immun-Star HRP Chemiluminescent Kit (BioRad, CA) and captured on Kodak X-Omat Film (Rochester, NY).

CHAPTER 2

SCREENING A LIBRARY OF DMSD BINDING POCKET MUTANTS FOR INCREASED AFFINITY TO DMSA WITH YFP-BiFC

2.1 Introduction

Bacterial protein evolution has been occurring over the last few billion years and proteins continue to evolve due to the selective pressures in the local environment. Evolution of these proteins at the genetic level occurs through in a variety of manners, a) random mutagenesis (e.g. UV exposure, DNA polymerase errors) leading to base pair insertions, deletions, transitions or transversions, b) through bacterial conjugation, c) transduction of genetic material by bacteriophage and d) transformation by uptake of DNA in the local environment (109). Within the last few decades, recombinant DNA technologies have emerged whereby the DNA coding for the protein of interest can be isolated, mutated, and subsequently the protein itself can be expressed (110). In this manner, proteins can be evolved *in vitro* at rates much higher than those seen in nature. One of the largest challenges faced by *in vitro* protein evolution is linking the phenotype (e.g. the activity of the protein) with the genotype (e.g. the proteins' DNA coding sequence).

To overcome this phenotype-genotype linkage problem, numerous reporter systems have been developed such as, a) ribosome display (111), b) phage display (112) and c) intrabody selection after Tat export (ISELATE) (113,114). Subsequently, large *in vitro* DNA libraries for specific proteins (e.g. single chain variable fragment (scFv) antibodies (113,115-118)) were generated and these reporter systems were applied to specifically detect proteins that had enhanced characteristics over the native protein of interest.

From our previous experimental results using the YFP-BiFC assay to monitor protein interactions *in vivo* in the *E. coli* Tat pathway, we observed that the fluorescent signal from YFP-BiFC interactions was dynamic and spanned an order of magnitude (e.g. DmsA-Y1/DmsD-Y2 versus TatA-Y1/TatA-Y2) depending on the type of protein interactions interrogated. Since the YFP-BiFC assay maintains a phenotype-genotype linkage through the expression of protein chimeras from a two plasmid system, cells with a specific phenotype can be recovered and their plasmids analyzed to detect genetic mutations that result in the observed phenotype. Therefore using this linkage we screened a plasmid library of DmsD mutants for an increase in YFP-BiFC signal that would arise from an increased affinity to ssDmsA.

2.2 Site directed mutagenesis of binding pocket residues in DmsD using an NNK primer library approach

To demonstrate the utility of our YFP-BiFC assay beyond screening natural components of the Tat system, we next sought to isolate gain-of-function DmsD variants that bind ssDmsA more efficiently, resulting in an increased YFP-BiFC signal. Previous studies utilizing *in silico* bioinformatic methods identified a "hot pocket" of residues which were conserved across numerous bacterial species possessing the homologous protein DmsD (Table 2.1). These "hot pocket" residues, referred to as binding pocket pt. 1 and pt. 2 (Table 2.1) were shown to be important for binding to the signal peptide of DmsA (1,119). Additionally the crystal structure of DmsD from *Salmonella typhimurium* was determined (Figure 2.1) allowing for an unprecedented view of the chaperone at a 1.38 Angstrom resolution (120). Using both of these resources, we went ahead and examined if we could screen a DmsA-Y1/DmsD-Y2 library using YFP-BiFC.

From published experimental data by Chan *et. al*. 2008 (1), two variants of

Table 2.1 Residues in DmsD involved in the binding of ssDmsA. Table adapted from Chan *et. al.* 2008 (1). After bioinformatic analysis, residues (x) were mutated using a site-directed kit and their ability to bind ssDmsA was interrogated with an *in vitro* dot-blot far-Western screen. Residues in bold showed a decrease in binding of ssDmsA compared to wt DmsD. Redox Enzyme Maturation Protein (REMP) familes, of which DmsD is a member, contain the two conserved residue motifs. The binding pocket in DmsD for ssDmsA consists of two parts, pt.1 (W72, L75, F76, P86) and pt. 2 (P124, D126, H127, L151).

	bioinformatics		
residue	REMP	DmsD	
G18		x	
Y22		x	
W72	x		Binding Pocket pt.1
L75	x	x	
F76	x	x	
P86		x	
W87		x	
D93		x	
E95		x	
E123	x	x	
P124	x	x	Binding Pocket pt.2
D126	x	x	
H127	x		
L151		x	
H154		x	
Y175		x	

DmsD were isolated; a) a lower affinity (compared to wt DmsD) double substitution mutant R15C/L75S and b) a hyperbinding variant of DmsD carrying a single amino acid substitution, W87Y, which was reported to bind ssDmsA at a level that was 1.5 times above that of wt DmsD when using a bacterial-two-hybrid assay (1). As a proof-of-concept, we created these mutant DmsD-Y2 chimeras (R15C/L75S and W87Y) for YFP-BiFC analysis, and co-expressed them with either ssDmsA-Y1 or DmsA-Y1 in wt TG1 cells. In agreement with the observations seen in Chan *et. al.*, the R15C/L75S mutant exhibited a decreased binding affinity to ssDmsA (i.e. a lower YFP-BiFC signal compared to wt DmsD), while the W87Y mutant had an increased affinity and in turn a higher YFP-BiFC signal compared to wt DmsD (Figure 2.2 A and B). Importantly, when these two DmsD mutants were coexpressed with full length DmsA, no fluorescence signal change was observed with YFP-BiFC when

compared to ssDmsA. This discrepancy can be explained initially by differences in experimental methods because both DmsD mutants (R15C/L75S and W87Y) were characterized for binding the signal peptide of DmsA, not full length DmsA. Additionally, the R15C/L75S mutation would have no affect on the YFP-BiFC signal because the full length DmsA-Y1 protein is 110kDa and undoubtedly interacts with other parts of the DmsD chaperone. These interactions result in a lower free energy state for the substrate-chaperone pair and counteract the disruptive point mutation.

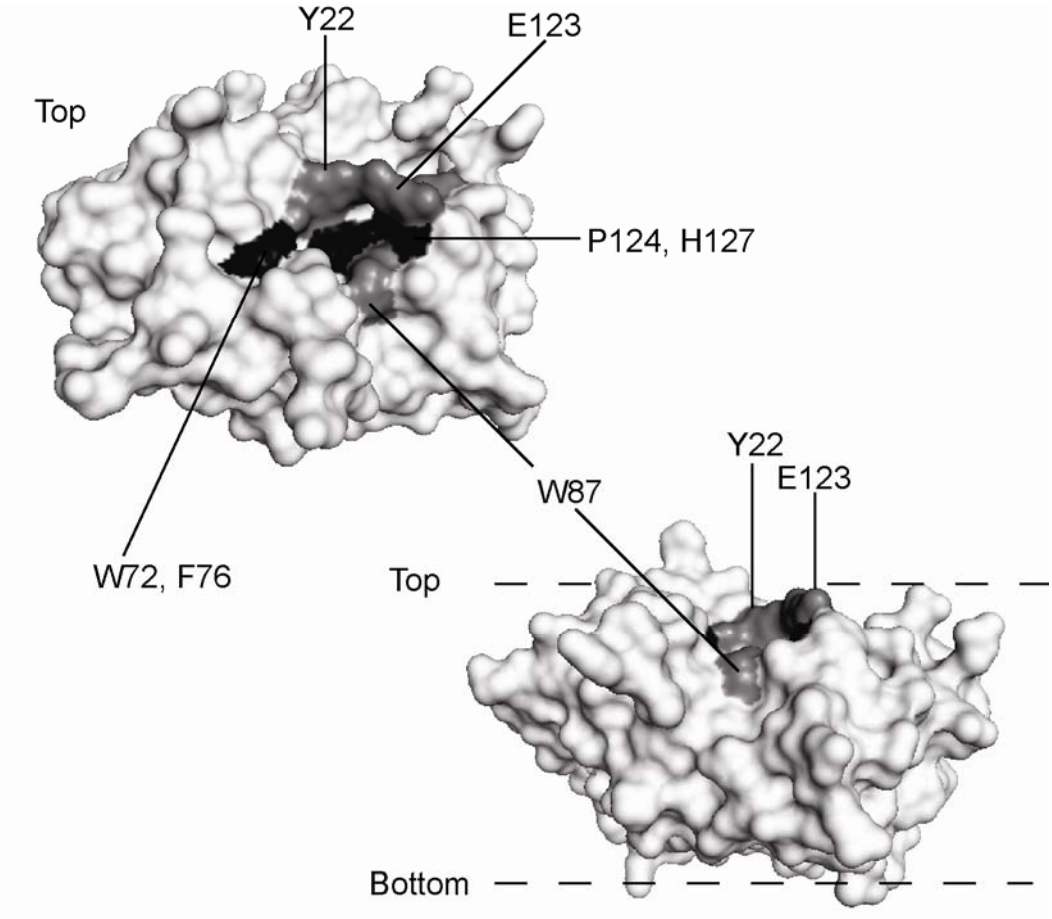

Figure 2.1 DmsD binding pocket for ssDmsA. Crystal structure of DmsD from *Salmonella typhimurium* by Qiu *et. al.* 2008 (2) (PDB ID 1s9u). Entire protein colored in white. Mutating residues colored in black (W76, F76, P124, H127) results in a decreased binding affinity to ssDmsA, while residues colored in gray (Y22, W87, E123) result in an increase in binding affinity for ssDmsA as demonstrated in Chan *et. al.* 2008 (1). Figure generated using PyMol (6).

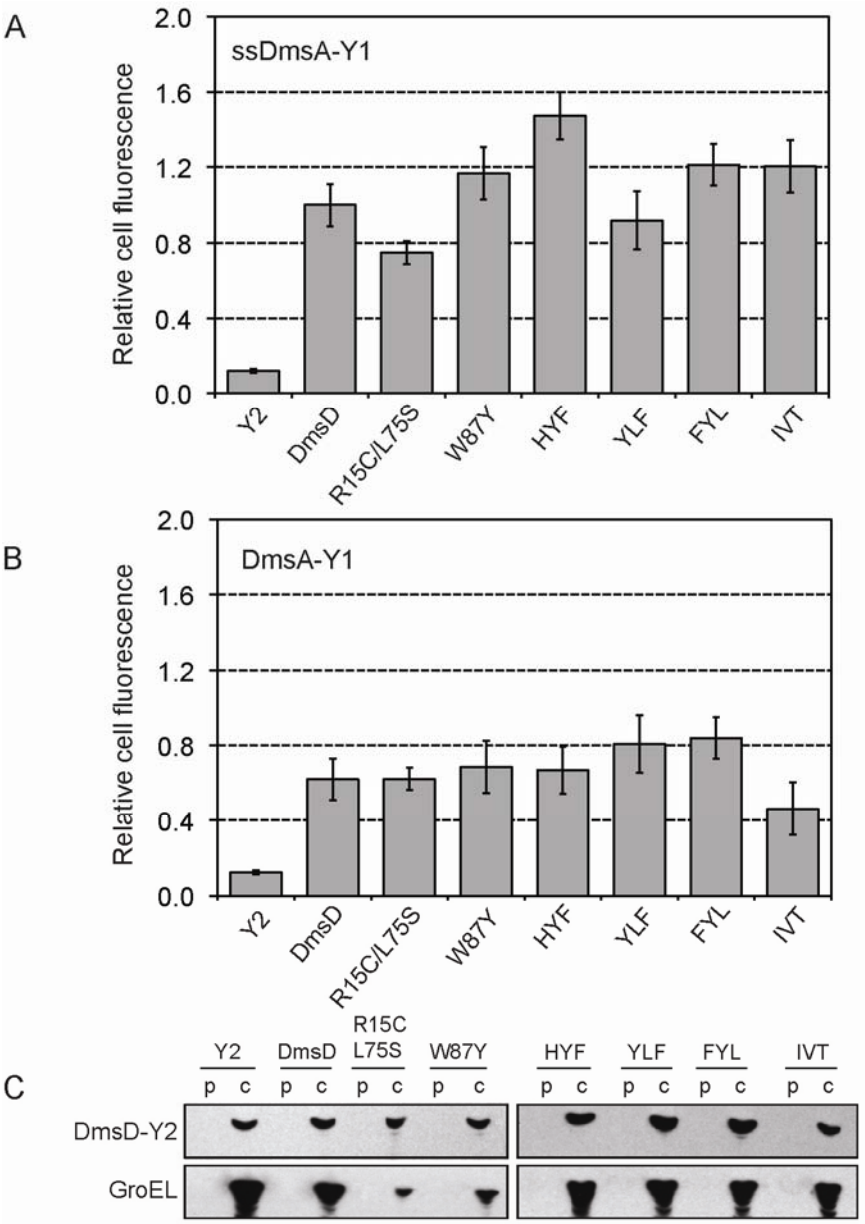

Figure 2.2 Isolation of gain-of-function chaperones. (A) Cell fluorescence of DmsD library isolates (HYF, YLF, FYL, IVT) when coexpressed with ssDmsA-Y1 or DmsA-Y1 (B) in wt TG1 cells. Two previously characterized mutants (R15C/L75S, W87Y) were included for comparison. Data normalized to ssDmsA-Y1/wt DmsD-Y2 signal and unfused Y2 coexpressed with ssDmsA-Y1 was the negative control. Median fluorescence values obtained via FACS and averaged from 3 replicate measurements ($n = 3$). (C) Western blot analysis of cytoplasmic (c) or periplasmic (p) fractions isolated from cells coexpressing ssDmsA-Y1 with DmsD-Y2 mutants as indicated. GroEL served as a cytoplasmic fractionation marker.

2.3 Screening a DmsD library with YFP-BiFC to identify clones with increased affinity for DmsA

Confident that our YFP-BiFC assay was able to discriminate changes in binding, we created 2 random libraries of DmsD using an NNK library approach (N coding: A,C,G,T; K coding: G,T) that targeted residues W72/L75/F76 in binding pocket pt. 1 (119) after analyzing the crystal structure with PyMol (6). The resulting DmsD-Y2 library was coexpressed with either ssDmsA-Y1 or full-length DmsA-Y1 in wt TG1 cells. After induction of protein expression for YFP-BiFC, the cells were screened via fluorescence activated cell sorting (FACS) for a fluorescent YFP-BiFC signal that was above the signal generated from wt DmsD-Y2/ssDmsA-Y1 or wt DmsD-Y2/DmsA-Y1 interactions.

The cells with higher YFP-BiFC signals were isolated and their plasmids containing the mutated variants of DmsD-Y2 were sequenced to identify the mutations generated by the NNK library. As seen in Table 2.2, a strong bias for hydrophobic, uncharged residues in these positions was observed, especially in positions 72 and 76

Table 2.2 Mutants isolated from 3-NNK DmsD library screen. The 3-amino acid variants isolated from the DmsD library are listed in the left column by their 3-letter designation. The DmsD-Y2 mutants were tested for YFP-BiFC with ssDmsA-Y1 or DmsA-Y1 and the fluorescence intensity was normalized to wt DmsD-Y2/ssDmsA-Y1 and reported as the binding activity for either ssDmsA or DmsA.

DmsD clone	Binding partner	Sequence	# of times isolated	ssDmsA binding activity	DmsA binding activity
wild-type	ssDmsA	71 AWQRLFV 77	1	1.00	0.62
HYF	ssDmsA	-H--YF-	2	1.48	0.67
YLF	ssDmsA	-Y--LF-	1	0.92	0.81
IVT	ssDmsA	-I--VT-	1	1.21	0.46
FYL	ssDmsA	-F--YL-	1	1.22	0.84
FDL	ssDmsA	-F--DL-	1	1.20	nd
FAP	ssDmsA	-F--AP-	1	0.90	nd
FQM	ssDmsA	-F--QM-	1	0.87	nd
VKM	ssDmsA	-V--KM-	1	1.09	nd
SNI	ssDmsA	-S--NI-	1	1.11	nd
SPH	ssDmsA	-S--PH-	1	1.09	nd
wild-type	DmsA	71-AWQRLFV-77	1	1.00	0.62
WMF	DmsA	-W--MF-	2	nd	0.63
WYF	DmsA	-W--YF-	2	nd	0.73
WFF	DmsA	-W--FF-	1	nd	0.59
FHL	DmsA	-F--HL-	1	nd	0.48
FHP	DmsA	-F--HP-	1	nd	0.46
FFP	DmsA	-F--FP-	1	nd	0.45

where a hydrophobic residue was found in 16/21 and 19/21 clones, respectively (7 and 10 of these, respectively, were wt in this position). Position 75 appears to be the most flexible as more than half of the clones carried a hydrophilic residue in this position, and in 2 of these cases the residue was charged (Arg, Asp).

Interestingly, these 3 residues (aa 72, 75, 76) of DmsD could tolerate much greater structural variability when the binding partner was ssDmsA-Y1 (Table 2.2). This indicates that substrate binding specificity is dependent on the context of the signal peptide (i.e. alone versus affixed to full length DmsA) and that the sequence determinants for binding of the full length DmsA are much more specific compared to just the ssDmsA signal peptide alone. Thus, care must be taken when interpreting data from chaperone binding experiments where signal peptides are used as surrogates for the full-length pre-protein substrate.

After generating and screening the NNK-DmsD library (as described in the Materials and Methods Section) we identified clones that had an increased YFP-BiFC fluorescent signal over the ssDmsA-Y1/DmsD-Y2 interaction where DmsD was wild type. From this screen we chose four mutants, named by their respective mutations (HYF, YLF, FYL and IVT) and coexpressed them with either ssDmsA-Y1 or DmsA-Y1 in wt TG1 cells. When the YFP-BiFC fluorescence data was normalized to the ssDmsA-Y1/wt DmsD-Y2 signal (Figure 2.2 A and B; Table 2.2), three of the mutants showed an increase in fluorescence (HYF, FYL, IVT) for the ssDmsA-Y1 chimera, while only mutants (YLF or FYL) demonstrated an increase in fluorescence for DmsA-Y1. To validate that the increased YFP-BiFC fluorescence signal correlated to improved substrate binding rather than simply higher protein expression levels for DmsD, we Western blotted against the DmsD-Y2 chimera and determined that the expression level of each clone was unchanged relative to wt DmsD (Figure 2.2 C).

From the characterized DmsD library isolates, we were able to obtain mutants

that showed a higher binding affinity to both the signal peptide and full length DmsA, validating that YFP-BiFC can be used as a tool to identify increased affinity mutants from a randomly generated library. However this library only focused on half of the residues that exist in the binding pocket of DmsD limiting the amount of sequence space that could contain higher affinity DmsD mutants.

2.4 Potential Future Library Screens

From the first NNK library screen, that targeted the three amino acids residues in the signal peptide binding pocket, we were able to demonstrate that enhanced variants of DmsD could be isolated with YFP-BiFC. Our first NNK library however was limited in scope because it only targeted half of the signal peptide binding pocket. Two sets of experiments are proposed to identify better binding DmsD variants using YFP-BiFC as a screening tool.

The first experiment will expand the NNK library to include all of the residues that are in the binding pocket, therefore increasing the sequence space of available amino acids that can be present and subsequently interact with the signal peptide of DmsA. The library size necessary to cover this sequence space is approximately $6.4x10^7$ clones, which is an acceptable and realistic size even for 3-fold coverage in *E. coli*. The selectivity of the YFP-BiFC assay will only detect those DmsD-Y2 clones that are soluble and can bind to either the ssDmsA-Y1 or DmsA-Y1 protein, resulting in a reconstituted YFP protein whose fluorescence would then be detectable by a FACS machine; thereby automatically eliminating non-binding DmsD variants from the library. DmsD clones isolated from this library can then be compared via DNA sequencing to determine if there are any consensus residues that result in increase of affinity to DmsA.

The second experiment is to quantify the relative affinities of the isolated

DmsD variants through surface plasmon resonance (SPR) experiments. The affinity for wt DmsD to ssDmsA has been determined by SPR (121), therefore the isolated DmsD mutants can be characterized in this manner to determine if there is a detectable increase in the relative affinity. This kinetic data could then be compared against the signal intensity observed *in vivo* for YFP-BiFC, and a correlation between affinity and signal intensity between the two experiments could be determined. This experiment would provide quantitative evidence that YFP-BiFC can be used as a tool to identify increased affinity protein interactions.

2.5 Discussion

Utilizing the previously characterized interactions between DmsA and DmsD from Chapter 1 using YFP-BiFC, we wished to expand the utility of YFP-BiFC to detect increased affinity DmsD variants to DmsA. As proof of concept experiments, we tested and confirmed previously characterized DmsD mutants with either decreased affinity (R15C/L75S) or increased affinity (W87Y) to ssDmsA using YFP-BiFC analysis. However, when we tested these same mutations against full length DmsA in the YFP-BiFC assay, we were not able to see an increase or decrease in signal intensity presumably due to interactions between sites outside of the binding pocket that stabilize the full length DmsA interaction with the DmsD chaperone.

With the proof of concept experiments validated, we created a 3-NNK library for amino acids located at positions 72, 75, and 76 in DmsD-Y2. This library was then screened against either ssDmsA-Y1 or DmsA-Y1 in TG1 cells. The cells were then sorted by FACS analysis and those DmsD-Y2 variants that exhibited a fluorescence YFP-BiFC signal over the wt DmsD-Y2 interactions were isolated and their DNA sequences were characterized. From this library screen we were able to identify numerous DmsD variants that showed an increase in YFP-BiFC for both the signal

peptide and full length DmsA protein.

From these experimental results, we have demonstrated for the first time that YFP-BiFC can be used as a tool to identify protein interactions with increased affinity.

2.6 Materials and Methods

2.6.1 Bacterial strains, growth and induction conditions.

Bacterial strains, growth, and induction conditions used in this study are described in Chapter 1.8, unless otherwise mentioned.

2.6.2 Construction of plasmids.

Previously used plasmids are described in Chapter 1.8, however additional plasmids used in this study are listed below in Table 2.3.

Table 2.3 Plasmids used in this study.

Plasmid	Description	Reference
pR15C/L75S	pDmsD-Y2 with R15C and L75S mutations	This work
pW87Y	pDmsD-Y2 with W87Y mutation	This work
pHYF	pDmsD-Y2 library isolate with W72H, L75Y, F76F mutations	This work
pYLF	pDmsD-Y2 library isolate with W72Y, L75L, F76F mutations	This work
pFYL	pDmsD-Y2 library isolate with W72F, L75Y, F76L mutations	This work
pIVT	pDmsD-Y2 library isolate with W72I, L75V, F76T mutations	This work

2.6.3 DmsD NNK library construction.

Bioinformatic and *in vivo* experimentation by Chan *et. al.* 2008 demonstrated that a potential "hot pocket" of residues involved in binding the signal peptide of *dmsA* occurred at two sites in the protein *dmsD*; site number one (amino acids W72, L75, F76), site number two (E123, P124, D126, H127). Using degenerate primers, MD1623 5'–GAA GAG ACT CAC GCC CAG GCC NNK CAG CGT NNK NNK

GTC GGC CCG TGG GCA CTG – 3' and MD1624 5' – CAG TGC CCA CGG GCC GAC CGC GCT ACG CTG GCT GGC CTG GGC GTG AGT CTC TTC – 3' (Integrated DNA Technologies, Inc., IA) whereby N (codes for A,C,G,T) and K (codes for G,T) with template plasmid pDmsD-Y2, we targeted the three residues in site number one with the Stratagene QuikChange II Site-Directed Mutagenesis Kit (La Jolla, CA). The amplified product was transformed into XL-1 Blue cells resulting in ~10^5 clones; representing a 3x library coverage. XL-1 Blue cells were harvested, grown in liquid culture and midi-prepped using the PureYield ™ Plasmid Midiprep System from Promega (Madison, WI).

The 3-NNK DmsD library was then digested with *Sph*I and *Kpn*I to isolate the mutated DmsD library and ligated into clean Kan^R plasmid backbone to prevent any potential mutations in the plasmid backbone that could be detrimental to the expression of the gene fusion. This second 3-NNK library was harvested and midi-prepped and then electroporated into electrocompetent TG1 cells which contained either pssDmsA-Y1 or pDmsA-Y1 using 2-mm gap electroporation cuvettes with preset electroporation settings (2.5kV, 25μF, 200Ω). This library was spread on LBA-Amp/Kan plates and incubated overnight at 37ºC for resolution of transformants.

2.6.4 Expression and isolation of enhanced affinity DmsD clones.

After overnight incubation, the transformed cells were pooled into a 500mL culture, allowed to grow to and OD_{600} ~ 0.5 at 37ºC and 200 rpm, induced with 1mM IPTG, and grown for 8hrs at RT/200 rpm. Aliquots of the culture were taken and diluted into one milliliter of 1x PBS and run through a BD FACSCalibur (Franklin Lakes, NJ) set for cell recovery mode. The gate used on the FACSCalibur would only recover cells with a fluorescent signal greater than the wt ssDmsA-Y1/DmsD-Y2 or the wt DmsA-Y1/DmsD-Y2 BiFC signal. The recovered cells were concentrated on Whatman Sterile Membrane Filters, 0.45μm pore size (Kent, UK) and the membrane

filter was transferred to LBA Amp/Kan plates to allow for single colonies growth overnight at 37ºC.

Isolates from the overnight incubation were then picked off and grown in 96-well culture plates to an OD_{600} ~0.5 at 37ºC and 200 rpm, induced with 1mM IPTG at RT/200 rpm for 8hrs. Fluorescence was then measured using a Biotek Synergy HT Plate Reader (Winooski, VT) with excitation filter 485/20, emission filter 528/20, bottom detector setting, and sensitivity of 50. Wells containing cells with a fluorescent signal greater than the wt ssDmsA-Y1/DmsD-Y2 or wt DmsA-Y1/DmsD-Y2 signal were then re-grown, plasmid DNA was harvested and submitted for sequencing. Selected sequences isolated from 96-well plates are listed in Table 2.2 while gain of function DmsD-Y2 characterized plasmids are listed above in Table 2.3.

2.6.5 Fluorescence measurements.

Cell fluorescence measurements were obtained via FACS as previously described in Chapter 1.8.

2.6.6 Cellular fractionation and Western blotting.

Cellular fractionation and Western blotting were performed as previously described in Chapter 1.8.

CHAPTER 3

UTILIZING YFP-BiFC TO STABILIZE PROTEIN INTERMEDIATES FOR PURIFICATION AND CRYSTALLOGRAPHY

3.1 Introduction

The ability to detect and monitor protein interactions *in vivo* is useful for understanding the macroscopic dynamics involved in intracellular systems. However if a more detailed study of the protein of interest is required (e.g. enzyme kinetics), *in vivo* methods become a hindrance due to the complex number of interactions occurring in parallel within the organism. Therefore it becomes a necessity to be able to isolate the protein of interest in a relatively short time span, to obtain a high purity sample of the protein without other protein contaminants, and most importantly perform these steps in an economically feasible manner for *in vitro* analysis.

The development of protein purification methods is a vast scientific area with numerous highly characterized methodologies and experimentally validated techniques, but it is also an area of continuous research and development. Affinity chromatography, whereby the protein of interest is selectively removed from cellular lysates due to a specific affinity for an immobilized substrate (e.g. a metal ion) is by far the most common method used for protein purification (122,123). Affinity chromatography methods are predominantly aimed at capturing a single protein of interest, but these methods can be modified to isolate protein complexes (57,124) that interact with the protein of interest; however some of these interactions are non-specific due to the *in vitro* nature of the purification environment.

The YFP-BiFC assay as described previously is able to monitor protein interactions occurring *in vivo*; however as mentioned before, the reconstitution of the

fluorescent protein YFP is irreversible under standard cellular conditions, resulting in the entrapment of the two interacting proteins. Utilizing the irreversible characteristic of YFP-BiFC, we demonstrate experimentally for the first time that the full length DmsA protein interacting with its chaperone DmsD can be purified to highly pure concentrations using traditional metal-affinity column chromatography methods.

3.2 YFP-BiFC – An irreversible association

The application of YFP-BiFC for the detection of protein-protein interactions *in vivo* is favored over other methods due to the assay's ability to directly monitor the interactions of two proteins in their local cellular environment and for the inability of the fragments to fold spontaneously, which would result in false positive signals (64,125,126). The folding of proteins is dictated by minimizing the free energy of the system (e.g. $-\Delta G_{\text{folded-unfolded state}}$); therefore folding of the two YFP fragments is energetically favorable, resulting in a stable-folded protein. This stability in turn ends up trapping of the two proteins of interest that are directly fused to the fragments and results in an irreversible association.

Since the YFP-BiFC assay is dependent on the affinity between the two proteins of interest, an interaction with a weak affinity will therefore not be detected. Using a leucine zipper library, the disassociation constant (K_D) of ~ 1mM was determined to be the affinity limit for the YFP-BiFC assay (70). Since the affinity between ssDmsA and DmsD is $K_D \sim 64\mu M$ (121), and a very strong fluorescent YFP-BiFC signal is observed for both the signal peptide and full length protein, we hypothesized that the full length DmsA-DmsD-YFP complex could be isolated.

3.3 Characterization of epitope tags for the purification of DmsA

To take advantage of the irreversible nature of the YFP-BiFC assay in

capturing protein interactions, we attached two common epitope tags to the C-terminus of full length DmsA (Figure 3.1). The FLAG (DYKDDDDK) (127) and His (HHHHHH) epitopes are highly immuno-reactive sequences that can be recognized by commercially available antibodies allowing for the quantitative analysis of recombinant protein production.

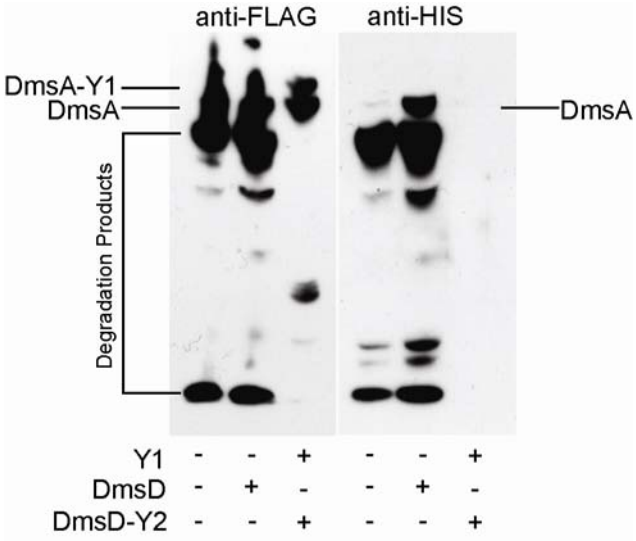

Figure 3.1 Using YFP-BiFC, substrate-chaperone pairs are stabilized and prevent proteolytic degradation. Two Western blots probed with primary-HRP conjugated antibodies against epitope tags FLAG or 6x-HIS, respectively. Either a FLAG tag or 6x-HIS repeat was cloned as a C-terminal fusion to full length DmsA or the YFP-BiFC chimeric DmsA-Y1. The DmsA constructs were expressed with either DmsD or DmsD-Y2 in Δ*tatC* TG1 cells. Only the soluble fraction was loaded for Western blot analysis.

Initially, we expressed full length DmsA with either a C-terminal FLAG or a HIS tag in Δ*tatC* TG1 cells to prevent the export of DmsA and cleavage of its signal peptide. In both expression studies we were able to detect full length DmsA, albeit at drastically different levels (FLAG versus HIS respectively). We additionally noted that a fair amount of degradation products were present, indicating that proteolytic degradation of the recombinant protein was occurring.

To alleviate the formation of degradation products, we coexpressed DmsD with DmsA fused to the two epitope tags at its C-terminus. Since DmsD is a

chaperone, coexpression of this protein would provide protection from premature proteolytic cleavage of the signal peptide and it also would assist in stabilizing DmsA while it attains its tertiary structure. Interestingly, we saw an increase in the amount of both FLAG and HIS DmsA that was produced, but unfortunately there was also an increase in degradation products evident on the Western blot.

Finally, we coexpressed DmsD-Y2 with either DmsA-Y1-FLAG or DmsA-Y1-HIS in Δ*tatC* TG1 cells, respectively. We were able to obtain two clean bands for the FLAG epitope with a drastically decreased amount of degradation products; however no expression was evident for the HIS construct. Expression of both DmsD-Y2 and DmsA-Y1-FLAG resulted in an irreversible association between the two proteins through YFP-BiFC and subsequently was able to resist degradation as had been seen for the previous cases.

3.4 Purification of the DmsA-DmsD YFP-BiFC complex for crystallography studies

In crystallography studies, the use of antibody fragments (e.g. Fab – fragment, antigen binding region) (128,129) or the fusion of T4-lysozyme (130) has been demonstrated to stabilize the protein complex and allow for the formation of crystals. In this same manner, we envision the irreversible nature of YFP-BiFC in helping to stabilize the protein complex and in turn aid in the formation of crystal structures between the DmsA-DmsD-YFP complexes.

Building upon the experimental results seen from the Western blot data in Figure 3.1, we engineered DmsD-Y2 to have an N-terminal 8x-HIS epitope, creating the plasmid, p8x-HIS-DmsD-Y2. This N-terminal HIS epitope was previously characterized (57) and shown to have no affect on the ability of DmsD to bind to ssDmsA. Additionally, the N-terminal HIS epitope allowed for the purification of DmsD from cell lysates using commercially available nickel-nitrilotriacetic acid (Ni-

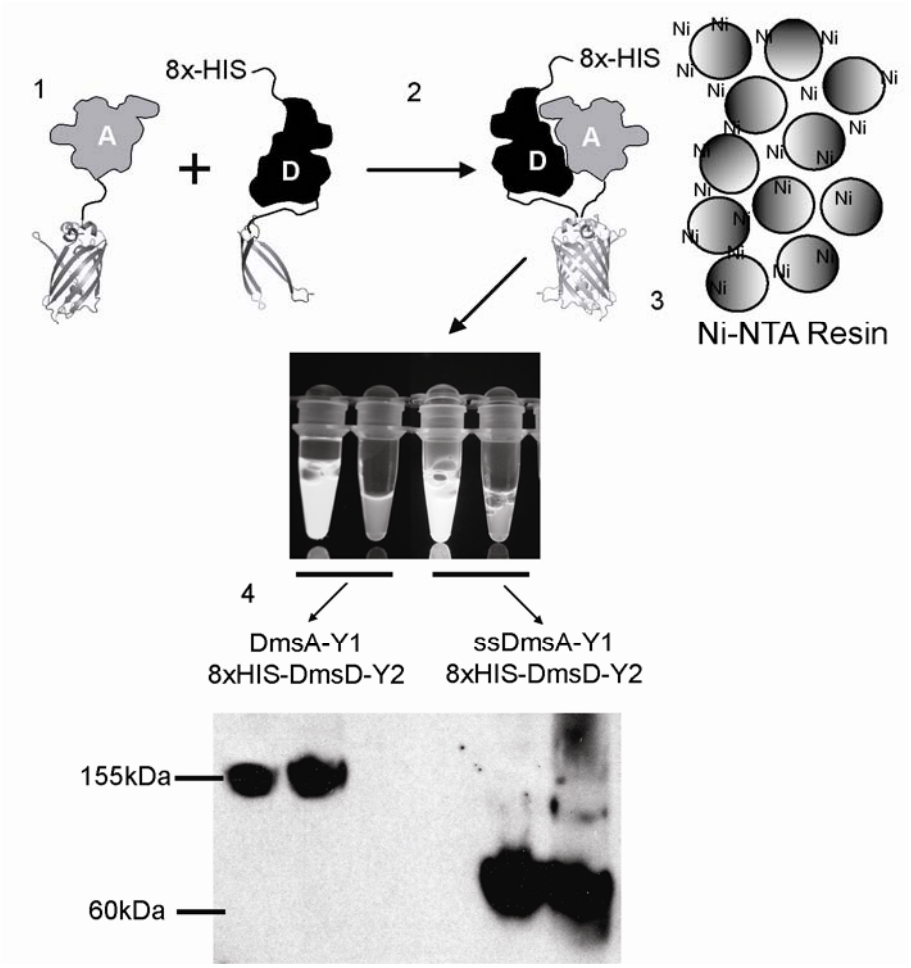

Figure 3.2 Purification of YFP-BiFC DmsA/DmsD complex. Flowchart depicting process for purifying YFP-BiFC protein complexes beginning with (1) coexpression of protein A (ssDmsA-Y1 or DmsA-Y1) with protein D (8x-HIS-DmsD-Y2). *E. coli* cells are lysed (2), the lysate is applied to a Ni-NTA resin for affinity purification (3). After protein samples are collected and concentrated, a native poly- acrylamide gel is run for the detection of the protein complex (DmsA/DmsD or ssDmsA/DmsD YFP-BiFC complexes are shown above).

NTA) purification columns. Coexpression of either ssDmsA-Y1-FLAG/8x-HIS-DmsD-Y2 or DmsA-Y1-FLAG/8x-HIS-DmsD-Y2 was preformed in Δ*tatC* TG1 cells and a fluorescent signal was detected by both FACS and fluorescence microscopy (data not shown).

After induction of protein expression, the bacterial cells were centrifuged, lysed, and then passed over a Ni-NTA column for metal affinity purification with the

8x-HIS tag attached to DmsD-Y2 (Figure 3.2). This column was then washed extensively to remove any non-specifically bound proteins and subsequently a high concentration of imidazole was applied resulting in the elution of the YFP-BiFC complex for either ssDmsA or DmsA. This flow through was collected and concentrated resulting in a highly fluorescent solution containing the ssDmsA/DmsD or DmsA/DmsD YFP-BiFC complexes.

The purified protein complexes stabilized by YFP-BiFC for either ssDmsA/DmsD or DmsA/DmsD were separated by native polyacrylamide gel electrophoresis (PAGE) and then transferred and detected by Western blotting against the C-terminal FLAG tag. As shown in Figure 3.2, two highly pure complexes were detected which are representative of the DmsA/DmsD and ssDmsA/DmsD interactions bound by the folded YFP from the YFP-BiFC interaction. To further validate that we had obtained the DmsA/DmsD YFP-BiFC complex, we applied the purified protein to a denaturing SDS-PAGE gel and obtained two distinct bands for DmsA-Y1-FLAG and 8x-HIS-DmsD-Y2 (Figure 3.3). Of importance, the folding of the YFP fragments does not result in a covalent bond forming, therefore by running a SDS-PAGE gel the

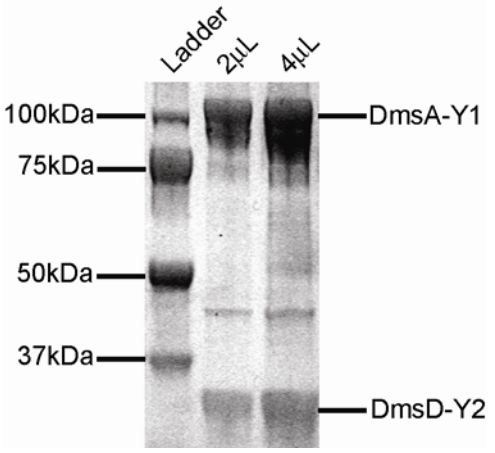

Figure 3.3 SDS-PAGE analysis of the YFP-BiFC DmsA/DmsD complex. (A) A 4-20% protein denaturing SDS-PAGE gel was run and stained with Comassie blue for visualization of the purified protein. Purified YFP-BiFC DmsA-Y1/8x-HIS-DmsD-Y2 was loaded into each well (2μL or 4μL of protein sample in 2x SDS-PAGE running buffer). Two major bands are apparent on the gel, one for DmsA-Y1 (MW~110kDa) and the second for 8xHIS-DmsD-Y2 (MW~34kDa).

two fragments will denature and separate from one another allowing for a spatial resolution of the two proteins. In addition to the two main bands (DmsA and DmsD) detected on the SDS-PAGE gel, a third band with a molecular weight of ~ 45kDa was also seen. The identity of this protein is unknown and could either associate with the DmsA-DmsD-YFP-BiFC complex or could just be a protein with a high affinity for the Ni-NTA purification column. If necessary, further analysis such as N-terminal protein sequencing and mass-spectroscopy would need to be performed to identity this protein.

Combining the technique of YFP-BiFC assembly with Ni-NTA affinity purification we were able to demonstrate that we can purify the two protein complexes, ssDmsA/DmsD or DmsA/DmsD which have been trapped in an energetically stable configuration through the assembly of YFP.

3.5 Discussion

Affinity purification of recombinantly expressed proteins using specially designed affinity tags is a well characterized field with extensively validated protocols and methodologies. However, not all recombinantly expressed proteins can tolerate the attachment of affinity tags and the stability of over-expressed protein products in cells can result in proteolytic cleavage and degradation.

Utilizing the irreversible nature of the YFP-BiFC assay, we demonstrate that the ssDmsA/DmsD and DmsA/DmsD complexes when held together by YFP are stable, they can be purified using Ni-NTA affinity column chromatography techniques, and can resist proteolytic degradation. Additionally, these YFP-BiFC complexes are relatively pure and predominantly contain the two interacting protein partners with some minimal protein contamination. Obtaining a relatively pure concentrated protein solution and also being able to form a stable protein structure are

the essential characteristics for protein X-ray crystallography studies. We are currently working with a collaborator, Ailong Ke at Cornell University to obtain the crystal structure of the YFP-BiFC DmsA/DmsD complex.

In conclusion, we have experimentally validated that the YFP-BiFC assay can be combined with Ni-NTA metal affinity chromatography to isolate highly stable and contaminate free protein complexes. We are also the first group to experimentally demonstrate that full length DmsA can purified in complex with its chaperone, DmsD.

3.6 Materials and Methods

3.6.1 Bacterial strains, growth and induction conditions.

Bacterial strains, growth, and induction conditions used in this study are described in Chapter 1.8, unless otherwise mentioned.

3.6.2 Construction of Plasmids

All plasmids used in this study are previously described in Chapter 1.8, with the exception of the plasmids listed below in Table 3.1. For pDmsA-FLAG, an internal *Kpn*I cut site was designed before the existing C-terminal FLAG epitope on the Y1 fragment. This extra *Kpn*I cut site allowed removal of the Y1 fragment and resulted in a direct DmsA-FLAG fusion. The pDmsA-Y1-6xHIS plasmid was created by PCR amplification of the Y1 fragment using primer MD1768, 5' – CTC ATG ATC GAT TTA ATG GTG ATG GTG ATG GTG GGC CAT GAT ATA GAC GTT GTG G – 3', containing the 6xHIS epitope, and then cloned into *Kpn*I/*Cla*I cut pDmsA-Y1 backbone. To create the pDmsD plasmid, DmsD was PCR amplified and cloned into pDmsD-Y2 digested with *Hind*III/*Cla*I, resulting in removal of the DmsD-Y2.

The p8x-HIS-DmsD-Y2 construct was created with forward PCR primer MD2395, 5' – CTC ATG TCT AGA AAT GCA CCA TCA CCA TCA CCA TCA CCA TAC CCA TTT TTC ACA GCA AGA – 3' and reverse primer MD3124, 5' –

CTC ATG ATG ATT CAT AAC TTT CTG TTT CGG TAC C – 3', to attach eight histidine residues to the N-terminus of DmsD. This PCR product was digested with *Xba*I and *Kpn*I and replaced wt DmsD in pDmsD-Y2.

3.6.3 Expression and purification of YFP-BiFC complex.

During the initial characterization studies of full length DmsA with C-terminal FLAG or HIS epitope tags, expression of the DmsA constructs was performed in Δ*tatC* TG1 cells with or without the coexpression of DmsD. The protein expression conditions for the constructs tested were the same as those during YFP-BiFC assays.

Table 3.1 Plasmids used in this study.

Plasmid	Description	Reference
pDmsA-FLAG	The Y1 fragment was removed from pDmsA-Y1 and a C-terminal FLAG epitope tag was attached	This study
pDmsA-6xHIS	The Y1 fragment was removed from pDmsA-Y1 and a C-terminal 6xHIS epitope tag was attached	This study
pDmsA-Y1-6xHIS	The C-terminal FLAG tag on pDmsA-Y1 was replaced with a 6xHIS epitope tag	This study
pDmsD	The Y2 fragment from pDmsD-Y2 was removed	This study
p8x-HIS-DmsD-Y2	8x-HIS tag added to N-terminus of DmsD, replacing wt DmsD in pDmsD-Y2.	This study

TG1 Δ*tatC* cells were co-transformed with either ssDmsA-Y1/8x-HIS-DmsD-Y2 or DmsA-Y1/8x-HIS-DmsD-Y2 and transformants were recovered as described in Chapter 1.8. Coexpression of the proteins was performed as described in Chapter 1.8, however cells recovered after transformation were first pre-cultured for 12 hours at RT/200 rpm in 5mL of LB with antibiotics added. The following day, 2mL of the pre-culture was added to 400mL of liquid LB with antibiotics and allowed to grow to an OD_{600} ~ 0.5 at 37ºC/200 rpm, induced with 1mM IPTG, and moved to RT/200 rpm for 8 hours for protein expression.

Cells were pelleted at 11,000 x g for 20 minutes at 4ºC. Supernatant was discarded and cell pellets were frozen at -20ºC. Cell pellets were thawed on ice and

the following compounds were added: 5mL of Novagen Bug Buster Master Mix per gram of wet cell pellet, the protease inhibitor phenylmethanesulphonyl fluoride (PMSF) at a final concentration of 5mM, and 10% (v/v) glycerol. Cells were then resuspended at RT on a 200 rpm rotating platform for 20 minutes and then the cell lysate was then spun at 11,000 x g for 20 minutes at 4ºC. From this point forward, the supernatant was kept chilled on ice, passed through a 0.22µm syringe filter to remove any whole cells or cellular debris from the soluble fraction. Imidazole was added to a final concentration of 15mM.

With the aid of a peristaltic pump, the protein solution was passed over a pre-equilibrated GE HisTrap HP 1mL column at 1mL/min flow rate. The column was washed with 100mL of binding buffer (500mM NaCl, 20mM Tris-HCl, 15mM imidazole, 10% glycerol, pH=7.9), 200mL of wash buffer (same as binding buffer but 80mM imidazole) and eluted with 25mL of elution buffer (same as binding buffer but 500mM imiddazole). The elution fraction was concentrated on 50kDa molecular weight cut off (MWCO) VivaSpin20 centrifugal concentrators and then placed into a 10kDA MWCO dialysis cassette. Dialysis of the protein was performed overnight at 4ºC into 1.5L of 25mM HEPES pH = 7.5, 75mM NaCl, 10% glycerol, and 5mM dithiothreitol (DTT).

3.6.4 Verification of epitope characterization studies and protein purification.

For the epitope characterization studies, equal amounts of cells (normalized by OD_{600} readings) were lysed with Bug Buster Master Mix and the provided protocol from the manufacturer for the isolation of soluble proteins was used. After mixing the soluble proteins in 2x SDS PAGE load dye with beta-mercaptoethanol and incubating for 15min at 100ºC, equal volumes of samples were loaded onto a 10% SDS-PAGE gel and run for 2 hours at 110 Volts (Figure 3.1).

For the purified proteins (ssDmsA-Y1 or DmsA-Y1 with 8x-HIS-DmsD-Y2),

after overnight dialysis, two 10μL aliquots of purified protein were removed. The first 10μL was added to 2x PAGE loading dye (4mL glycerol, 1mL 0.1% bromophenol blue, 2.5mL 0.5M Tris-HCl pH = 6.8), mixed, and applied to a 4-20% PAGE gel, run at 100 Volts for 120 minutes resulting in native-non-denatured protein separation (Figure 3.2). The other 10μL of purified protein was mixed with 2x SDS-PAGE loading dye (same as 2x PAGE dye but with 0.5mL of beta-mercaptoethanol and 0.4g of sodium dodecyl sulfate) and incubated at 100ºC for 15 minutes for denaturing gel analysis. Either 2μL or 4μL of sample in loading dye was applied to a 4-20% SDS-PAGE gel and run for 120 minutes at 100 Volts (Figure 3.3 for just DmsA-Y1/8x-HIS-DmsD-Y2). The gel was then stained with BioRad BioSafe Coomassie Blue and the image was captured on a BioRad ChemiDoc.

3.6.5 Western blotting of epitope characterization studies and of purified protein.

The SDS-PAGE gel for the epitope characterization studies was transferred onto PVDF membranes and probed against the FLAG or HIS epitope tags using previously described antibodies for FLAG, while the HIS epitope was detected with a primary antibody against HIS conjugated to HRP. The native protein gel was transferred onto PVDF membranes and probed against the FLAG epitope as described in Chapter 1.8.

CHAPTER 4

UTILIZING YFP-BiFC TO ANALYZE PROTEIN INTERACTIONS BETWEEN A DE NOVO DESIGNED THREE HELIX BUNDLE PROTEIN AND THE E. COLI TAT PATHWAY

4.1 Introduction

The two predominant protein transport pathways in bacteria, the General Secretory pathway (Sec) and Signal Recognition Particle (SRP) pathway, allow for the transport of proteins into the periplasm in an unfolded or in a co-translational manner, respectively (9,131). In contrast, the Tat pathway is able to export fully folded proteins and protein-complexes (e.g. the hitch-hiker mechanism) through the inner membrane utilizing the proton motive force that exists between the periplasm and cytoplasm (54,78,132,133). The Tat pathway however does not transport unfolded proteins (134,135) due to its protein proofreading and quality control abilities.

The Tat translocon (e.g. TatABC) itself acts as a protein proofreading mechanism by inhibiting proteins that are incompetent for transport into the periplasm. The most basic level of proofreading exists at the recognition of the twin arginine residues, whereby the mutation of these residues to twin-lysines allows for the protein to interact with and recognition by the Tat machinery components, but it is subsequently prevented from being transported into the periplasm (136,137). The Tat pathway chaperone DmsD serves as a quality control mechanism by stabilizing the folding of DmsA and by preventing premature export of the substrate before cofactor insertion (57,138). These two methods provide the Tat pathway with a degree of specificity and selectivity as to the types of proteins that can be transported into the periplasm.

From a biotechnological application perspective, the ability to have an *in vivo* proofreading and quality control mechanism would provide for an internal checkpoint for the production of high value therapeutics without the need to worry about the folding state of the protein of interest. Understanding how the Tat pathway is able to recognize and prevent export of unfolded or export-incompetent proteins has been investigated (90,134,139), and a selection tool utilizing the export of the antibiotic resistance conferring protein beta-lactamase has been designed to identify proteins that are Tat export competent from a *de novo* designed library (114); however a direct *in vivo* protein reporter that allows interrogation of the Tat pathway components with the *de novo* designed protein has yet to been developed.

Here we demonstrate the utility of the YFP-BiFC assay and its ability to report on the folding status of a *de novo* designed three helix bundle protein as it interacts with either the TatC or DmsD.

4.2 Description of α_3A – α_3B de novo designed proteins

Within the last decade, especially after the sequencing of the human genome, the number of fully sequenced genomes across a broad range of organisms has continuously increased. Additionally, the protein crystallography community has expanded the number of solved protein crystal structures with advancements in high throughput automated methods for the production, purification, and screening of proteins of interest for potential drug therapies (140,141). The information obtained from both the genomic DNA sequences and protein structures has allowed for the development of computer simulations that can begin to predict the function of unknown proteins (142).

Knowledge gained from the identification and characterization of native proteins has provided the field of *de novo* protein design with a framework for the

necessary conditions that allow for proteins to attain a stable folded structure and perform a certain function (143,144). The initial starting point for *de novo* protein design is done with bioinformatic analysis and *in silico* modeling of small protein sequences (consisting of 30-70 amino acid residues) that are patterned after naturally occurring motifs such as zinc fingers (145-148), coiled coils (149), or other small protein domains (150,151). Typically these proteins have a secondary structure consisting of alpha helices that can subsequently interact with one another to form bundles.

For our proof of concept experiments we wished to use the YFP-BiFC assay as a reporter on the folding state of a *de novo* designed protein as it interacts with either TatC or DmsD. Therefore we chose a *de novo* three helix bundle protein (3,4) that had a clear directed evolutionary lineage, was well characterized, and could be easily manipulated for the YFP-BiFC assay. The three helix bundle family of proteins we selected named, α_3A through α_3D are shown below (Table 4.1) and their respective amino acid sequence alignment is shown in Figure 4.1.

Table 4.1 Characteristics of α_3 helix bundle proteins. Proteins α_3A to α_3D, described below, were evolved through bioinformatic and computational analysis from the parental 2 helix bundle Coil-Ser protein as described in Bryson *et. al.* 1998 (4).

Peptide	Structure	Features
α_3A	Molten Globular	Parental platform: Coil-Ser protein - 2 helix bundle Hairpin loops added between helices - failure to stabilize the bundles Aggregation prone; monomer/dimer/trimer structures form
α_3B	Molten Globular	Addition of helix capping boxes - strong start/stop signal for helix Able to obtain a defined 3 helix bundle Minimzed aggregation; predominatly monomeric structures Possesed molten globular characteristics Lacked a nonpolar core which is essential for functional proteins
α_3C	Ordered	Repacked and rearranged the hydrophobic core Near "native like" - heat capacity, denaturation via GdnHCl, ANS binding Monomeric
α_3D	Ordered	Diversification of amino acids (19 of 20 aa used) Native like characteristics compared to natural proteins Monomeric

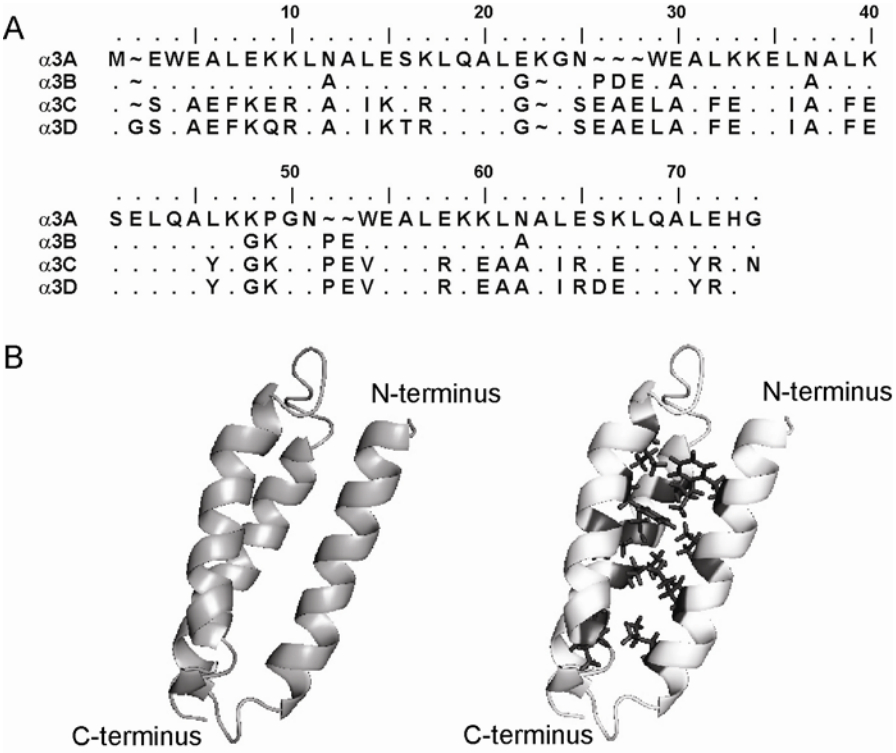

Figure 4.1 Description of *de novo* designed three helix bundle proteins α_3A-α_3D. (A) Protein alignment of α_3A-α_3D using ClustalW in DNASTAR LaserGene software package, with α_3A as base for protein evolution. Symbol (.) is a conserved residue, while (~) is a amino acid insertion. (B) NMR structure of the three helix bundle protein α3D as described by Walsh *et. al.* 1999 (3) (PDB ID 2a3d) displayed without or with internal hydrophobic amino acid residues (left and right images, respectively).

The α_3 helix bundle family of proteins originates from a *de novo* designed two helix bundle protein, Coil-Ser (152) that was created with a repeating periodic structure and was intended to form a homodimer that would stabilize the two bundles. When analyzed in solution, the peptide existed as a combination of monomer, dimers, and trimers while its X-ray structure showed an anti-parallel three-stranded coiled coil (153) demonstrating that the anticipated *de novo* designed structure was invalid and needed to be refined. Through directed evolution techniques based off of the determined X-ray structure, α_3A was created whereby hairpin loops were engineered to stabilize the bundles, but the protein still remained molten globular and was still aggregation prone upon analysis.

Further protein evolution generated α_3B, which was now a monomeric protein due to the addition of specific helix-capping motifs to the loop regions of the bundles, however this protein still was classified as having a molten globular structure due to the lack of hydrophobic residues at the center of the protein. Through the combined efforts of redesigning the electrostatic interactions between the helices and utilizing *in silico* modeling algorithms to repack and generate a hydrophobic core, the α_3C evolved variant was created. When analyzed, α_3C had an ordered monomeric three helix bundle structure and behaved like a native protein when denatured and tested for the lack of binding to 8-anilino-naphthalene sulfonic acid (ANS) which is a fluorescence dye that can bind to molten globular proteins (4).

Finally, the α_3D protein was evolved from α_3C by expanding the amino acid diversity to include 19 out of the 20 natural amino acids and when characterized, α_3D retained a monomeric and ordered structure. Furthermore the α_3D protein was stable enough for nuclear magnetic resonance (NMR) solution structure solving (Figure 4.1 B) and demonstrated the characteristics that were anticipated from the *de novo* design parameters (3). Having this family of three helix bundle proteins, which have been engineered for their stability, their ability to exist as monomers in solution and their native like resemblance of natural proteins, we were now able to utilize the YFP-BiFC assay to determine how dynamic interactions between folded and unfolded proteins occurs between Tat pathway components.

4.2 Interactions between three helix bundle proteins and DmsD monitored with YFP-BiFC

The first native Tat pathway interaction we studied and validated as a proof of concept with YFP-BiFC was the association between ssDmsA-Y1 and DmsD-Y2. Utilizing the affinity of the chaperone DmsD for ssDmsA, we created a tripartite

fusion protein with an N-terminal ssDmsA signal peptide, one of the four α_3 helix proteins, and the C-terminal Y1 fragment (e.g. ssDmsA-α_3A-Y1) and coexpressed these with DmsD-Y2 in wt TG1 cells for YFP-BiFC analysis (Figure 4.2 A).

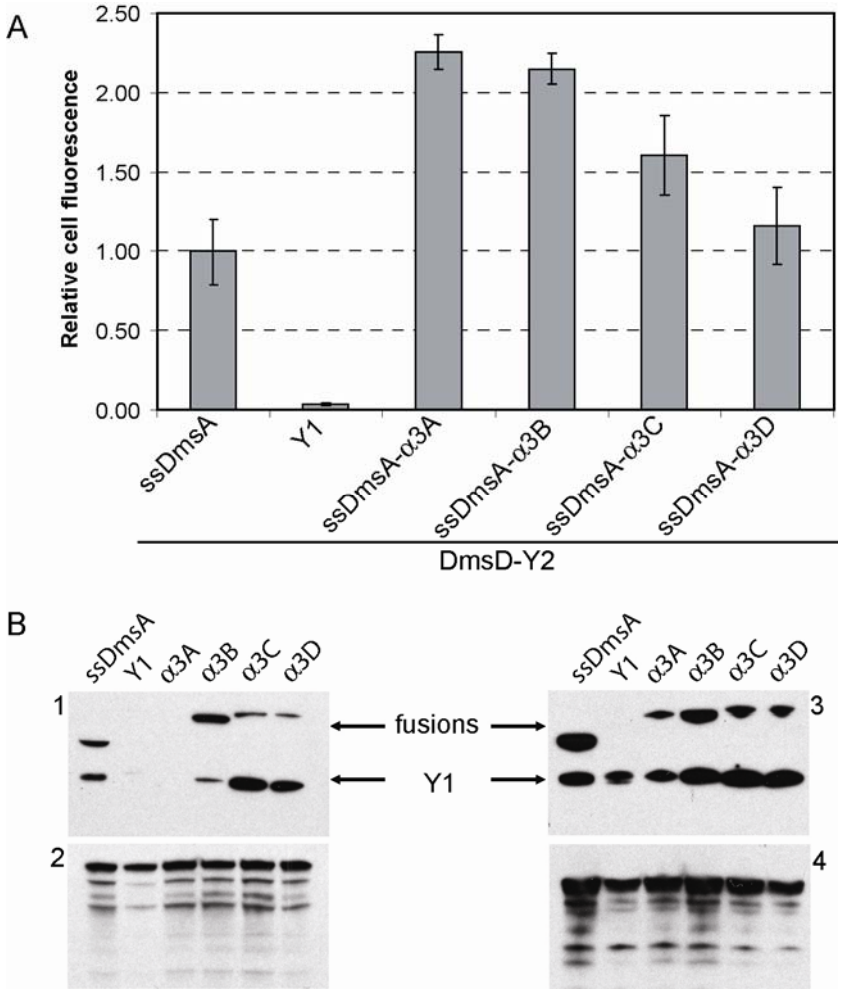

Figure 4.2 YFP-BiFC analyses of interactions between ssDmsA-α3-Y1 and DmsD-Y2. (A) Coexpression of ssDmsA-Y1, ssDmsA-α3A-Y1 through ssDmsA-α3D, with DmsD-Y2 in wt TG1 cells. Unfused Y1, was coexpressed with DmsD-Y2 and served as the negative control. Median cell fluorescence was obtained via FACS with 3 replicate experiments (n = 3) and was normalized to the ssDmsA-Y1/DmsD-Y2 signal. (B) Western blot analysis of periplasmic (panel 1 and 2) and cytoplasmic (panel 3 and 4) fractions from wt TG1 cells. All samples were fused to Y1 C-terminally, α_3A-α_3D had N-terminal ssDmsA fusions, while (-) was just Y1 and were coexpressed with DmsD-Y2. Panels 1 and 3 were detected using a anti-FLAG antibody against the C-terminal FLAG epitope on the Y1 fragment. Panel 2 is probed against GroEL, a cytoplasmic fractionation marker, with anti-GroEL. Panel 4 was probed with anti-GFP which detects the Y2 fragment fused to DmsD.

When the YFP-BiFC signal was normalized to the ssDmsA-Y1/DmsD-Y2 interaction, a greater than 2 fold signal was observed for the protein chimeras with α_3A and α_3B and slowly began to decrease for α_3C, while the α_3D signal was near that seen for the ssDmsA interaction. This decrease in fluorescence can be attributed to the flexibility and folding state of the α_3 helix bundle that is fused between ssDmsA and Y1. We hypothesize the relative steric hindrances for molten globular structures (e.g. α3A) are much lower than that of an ordered protein (e.g. α_3D), which can increase the probability that the two YFP fragments will interact with one another and this is reflected with an increase in detectable YFP-BiFC signal.

After seeing this decrease in fluorescence levels as the stability of the α_3 helix protein increased, we wished to validate that the decrease in signal was not due to a lower expression level of either the ssDmsA chimeric protein or the chaperone DmsD. The relative levels of expressed protein by visual inspection were similar when we fractionated the cells and Western blotted to detect either the ssDmsA chimeric protein or the chaperone DmsD in the cytoplasmic fraction (Figure 4.2 B, panel 3 or panel 4, respectively). Additionally, we looked for the transport of the ssDmsA-Y1 or ssDmsA- α_3A/α_3B/α_3C/α_3D-Y1 chimeric proteins into the periplasm (Figure 4.2 B, panel 1) and we were able to detect the active transport of all the Tat competent substrates (ssDmsA-Y1, ssDmsA-α_3B/α_3C/α_3D-Y1), while the unfused Y1 protein and the ssDmsA-α_3A-Y1 were not present. In the case of the molten globular and aggregation prone α_3A protein, the chimeric protein can interact with the soluble cytoplasmic DmsD chaperone and generate an intense YFP-BiFC signal; however periplasmic export of this protein is prevented. A similar increase in YFP-BiFC signal was seen when Tat machinery components were eliminated (e.g. Δ*tatC* TG1 cells) and ssDmsA-Y1/DmsD-Y2 were coexpressed (Figure 1.7 A).

The YFP-BiFC assay demonstrates that protein interactions between a Tat

pathway chaperone and a family of *de novo* designed proteins can be monitored and that the fluorescent signal intensity is an *in vivo* reporter for the folding status of the protein of interest between the ssDmsA and Y1. Furthermore, Western blot evidence demonstrates that the quality control and proofreading mechanisms of the Tat pathway are functional and prevent export of Tat incompetent substrates. This is the first demonstration that the YFP-BiFC assay can detect interactions between either globular unfolded proteins or structured proteins with DmsD in a functional Tat pathway.

4.3 Interactions between three helix bundle proteins and TatC monitored with YFP-BiFC

After validating that the YFP-BiFC assay can report on and discriminate between the interactions of differentially folded proteins (e.g. the α_3 helix bundle family) with the Tat pathway chaperone DmsD, we wished to determine if YFP-BiFC assay could be applied to monitor differentially folded protein interactions with the Tat translocase component, TatC. Utilizing the previously designed TatC-Y2 chimeric protein we coexpressed the ssDmsA-α_3A/α_3B/α_3C/α_3D-Y1 in Δ*tatABCE* TG1 cells and detected a fluorescent YFP-BiFC signal which was normalized to the ssDmsA-Y1/TatC-Y2 control (Figure 4.3).

As previously seen for the chaperone DmsD case, the trend for the decrease in the fluorescent YFP-BiFC signal occurs as the α_3 helix bundle family moves from the molten globular structure to an ordered one (e.g. α_3A to α_3D) when it interacts with the TatC-Y2 chimeric protein. The low YFP-BiFC signal for the α_3A chimeric protein can be attributed to the aggregation prone nature of this protein which can prevent it from interacting with TatC. As for α_3B, an increase in YFP-BiFC fluorescence can be attributed to the flexibility of the chimeric protein whereby the steric hindrances are

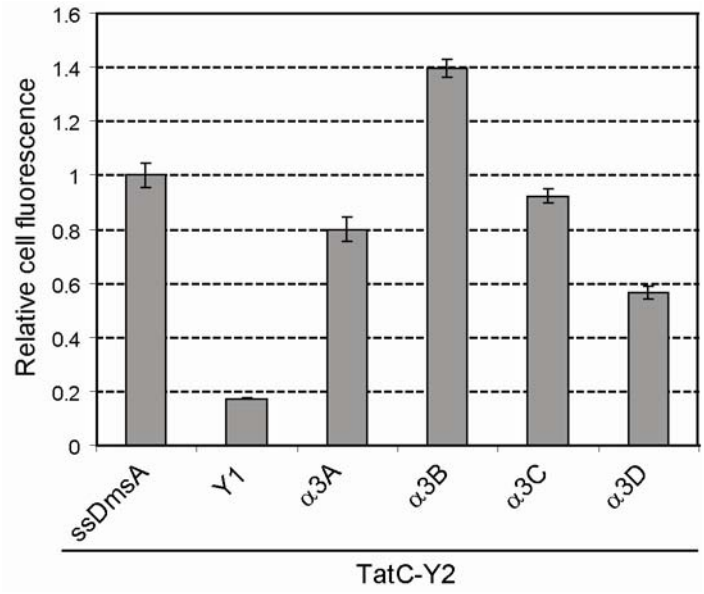

Figure 4.3 Interaction between TatC and ssDmsA-α3A-α3D-Y1 detected by YFP-BiFC. Proteins ssDmsA-Y1, ssDmsA-α3A-Y1 through ssDmsA-α3D-Y1 were coexpressed with TatC-Y2 in TG1 Δ*tatABCE* cells. Unfused Y1 was coexpressed with TatC-Y2 was used as a negative control. Median fluorescence values were obtained via FACS and the data was normalized to the ssDmsA-Y1/TatC-Y2 signal. The values reported are the average of 3 replicated measurements (n = 3).

minimized allowing for a greater number of YFP fragments to associate with one another. The α_3C chimeric protein YFP-BiFC signal resembled that of ssDmsA-Y1 while the α_3D chimeric proteins show a marked decrease in YFP-BiFC fluorescence that can be attributed the steric restrictions that the helix bundle imposes on the Y1 fragment and its ability to interact with TatC. Overall, we are able to demonstrate that the YFP-BiFC assay can be used to detect the interactions between differentially folded types of helical bundle proteins as they interact with a Tat translocon component.

4.4 Discussion

The coupling of *de novo* designed three helix bundle proteins of varying folded conformational states with the YFP-BiFC assay has allowed us to monitor the *in vivo* protein interactions between the α_3 chimeras and DmsD or TatC. By using a tripartite

fusion protein (e.g. ssDmsA-α_3A-Y1) coexpressed with DmsD-Y2, we were able to detect a fluorescent YFP-BiFC signal regardless of the folded state of the α_3 protein. However when we analyzed the ability of the Tat pathway to export these chimeric fusions by Western blotting, we were able to see the quality control and proofreading mechanism was fully functional and prevented the export of the ssDmsA-α_3A-Y1 fusion due to its molten globular and aggregation prone nature.

Additionally, we were able to detect interactions using YFP-BiFC between the same chimeric α_3 fusions and TatC-Y2 demonstrating that even unfolded, molten globular proteins can interact with the Tat translocon machinery, albeit at lower YFP-BiFC signal intensities. Although these experiments with TatC-Y2 were performed in Δ*tatABCE* TG1 cells, future experiments will be conducted in Δ*tatC* TG1 cells to determine how the YFP-BiFC signal intensity changes when other Tat components present.

Finally, we have demonstrated for the first time that the folding status of a *de novo* designed three helix bundle protein can be monitored via YFP-BiFC as it interacts with either a Tat chaperone or a Tat translocon machinery component. This fluorescent assay can be potentially applied to screen libraries of other *de novo* designed helix bundles (154) for the isolation and identification of other types of proteins that can interact with and be transported by the Tat pathway.

4.5 Materials and Methods

4.5.1 Bacterial strains, growth and induction conditions.

Bacterial strains, growth, and induction conditions used in this study are described in Chapter 1.8, unless otherwise mentioned.

4.5.2 Construction of plasmids.

All plasmids used in this study are previously described in Chapter 1.8, with

the exception of the plasmids listed below in Table 4.2. Plasmid pssDmsA-NdeI-SpeI-Y1 was designed by PCR amplification of ssDmsA with forward primer MD1007 5' – CTC ATG GCA TGC CTG CAG TAT GAA AAC GAA AAT CCC TGA TGC G – 3' and reverse primer MD 3051 5' – CTC ATG GGT ACC GCA CTA GTG TCG ACC ATA TGG GCG CTA TCG ACA GCG TGC GCA ATC CGA – 3'. This PCR product was digested with *Hind*III and *Kpn*I and inserted into pssDmsA-Y1 replacing the ssDmsA sequence. Plasmids containing the genes encoding for proteins α3A-α3D were provided as a gift of Dr. William F. DeGrado and were digested with *Nde*I and *Spe*I and cloned into similarly digested pssDmsA-NdeI-SpeI-Y1.

Table 4.2 Plasmids used in this study.

Plasmid	Description	Reference
pssDmsA-NdeI-SpeI-Y1	*Nde*I and *Spe*I cutsites were inserted after the ssDmsA signal peptide and before the *Kpn*I cutsite.	This work
pssDmsA-α3A-Y1	The α3A gene was inserted into the *Nde*I and *Spe*I sites of pssDmsA-NdeI-SpeI-Y1	This work
pssDmsA-α3B-Y1	The α3B gene was inserted into the *Nde*I and *Spe*I sites of pssDmsA-NdeI-SpeI-Y1	This work
pssDmsA-α3C-Y1	The α3C gene was inserted into the *Nde*I and *Spe*I sites of pssDmsA-NdeI-SpeI-Y1	This work
pssDmsA-α3D-Y1	The α3D gene was inserted into the *Nde*I and *Spe*I sites of pssDmsA-NdeI-SpeI-Y1	This work

4.5.3 Fluorescence measurements

Cell fluorescence measurements were obtained via FACS as previously described in Chapter 1.8.

4.5.4 Cellular fractionation and Western blotting.

Cellular fractionation and Western blotting were performed as previously described in Chapter 1.8.

CHAPTER 5

EXPANDING YFP-BiFC FOR FRET ANALYSIS OF PROTEIN INTERACTIONS ALONG THE TAT PATHWAY

5.1 Introduction

In this post genomic era, identifying the DNA coding for proteins has become trivial, however understanding the function of the protein of interest (POI) and subsequently mapping the interactions between the POI and other proteins that are expressed has been delegated to field of proteomics. The ability to identify protein-protein interactions can be preformed using *in vitro* methodologies such as affinity column chromatography whereby the POI is captured by an affinity tag to the column and then a cellular lysate is flowed over the column resulting in the association of proteins to the POI (57). After subsequent washing steps and elution of the POI-protein complex, the resulting proteins can be identified through N-terminal protein sequencing using Edman degradation (155) and/or mass spectroscopy analysis (156). This *in vitro* method however can result in a high number of false positive proteins due to non-specific interactions between proteins on the chromatography column.

The development of identification technologies that combine *in vivo* with subsequent *in vitro* analysis such as formaldehyde cross-linking with immuno-precipitation (157,158) begin to identify specific complexes, but again are limited due to the potential cross-linking effects by incorporating non-specific proteins. Yeast two hybrid (159) and bacterial two hybrid (160) systems are purely *in vivo* methods for the detection of protein-protein interactions whereby the interaction of two proteins causes transcriptional activation of a reporter gene and subsequent production of a reporter protein (e.g. beta-galactosidase or GFP). This *in vivo* method does not succumb to

non-specific protein interactions as the previously described methods do, however the assay gives an indirect report of the interaction of the two proteins since it is dependent on the transcription of a reporter protein.

Two purely *in vivo* methods, Förster/fluorescence resonance energy transfer (FRET) and bimolecular fluorescence complementation (BiFC), have been developed for the detection of protein interactions. Both of these methods utilize a fluorescent protein (e.g. CFP) or protein fragments (e.g. YFP) which are fused genetically to the POI and subsequent interactions can be monitored both temporally and spatially. Here we demonstrate that interactions between multiple Tat pathway proteins can be monitored using a CFP–YFP-BiFC–FRET assay.

5.2 The CFP–YFP-BiFC–FRET premise

We have previously demonstrated that YFP-BiFC can monitor protein-protein interactions in the *E. coli* Tat pathway; however this assay is limited in scope because it is only able to report on the interactions between two specific proteins (e.g. DmsA and DmsD). Since numerous proteins are involved in the Tat translocon, we wished to apply the technique of FRET to determine how these protein complexes interact with one another. FRET is a biophysical process whereby non-radiative energy is transferred from a donor molecule to an acceptor molecule as long as the two molecules are located within a distance of 10 nano-meters (nm) and their transition dipole moments are aligned (161,162).

The two most commonly used GFP spectral variants for FRET analysis are cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) (163). When CFP (ex. 434nm/em. 477nm) and YFP (ex. 513/em. 527nm) are farther than 10nm from one another, FRET will not occur; however when the distance is less than 10nm, the emission energy from CFP will excite YFP and an increase in the relative YFP to

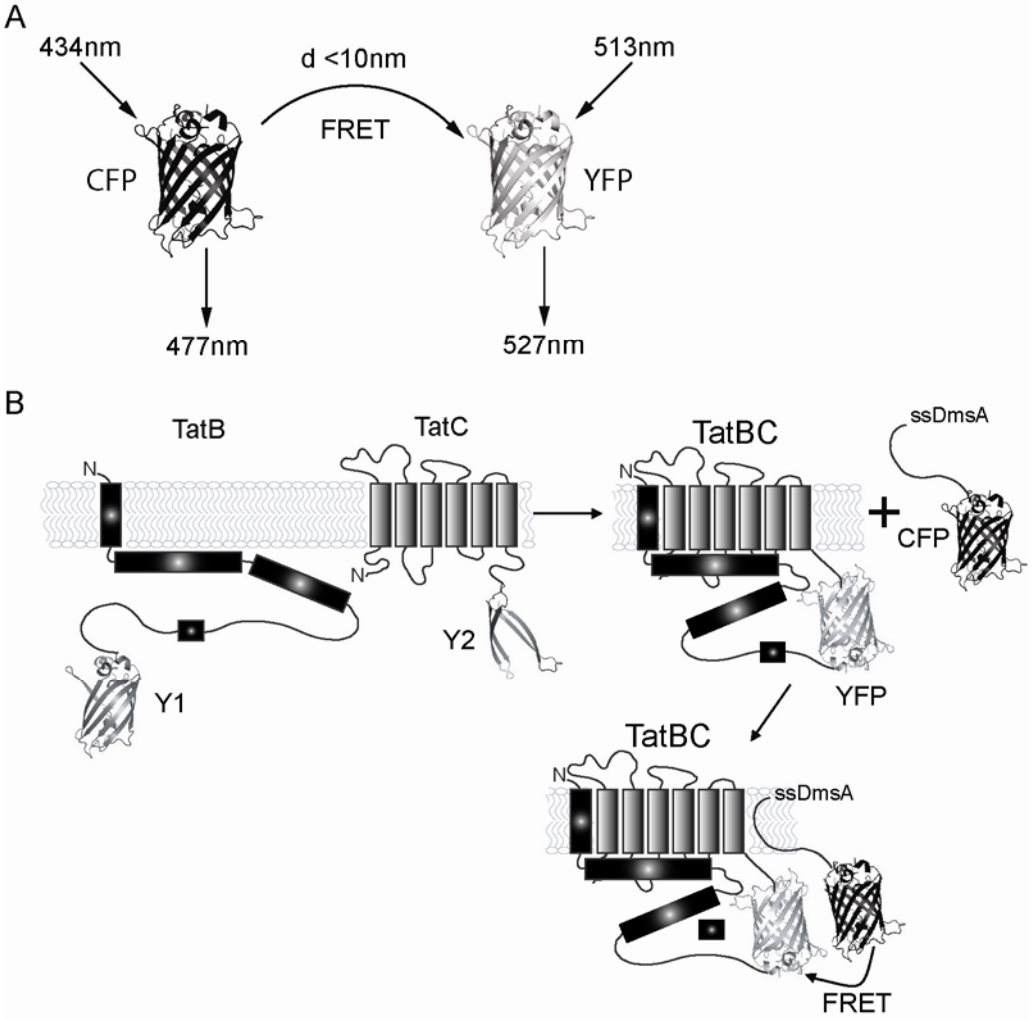

Figure 5.1 Applying YFP-BiFC-FRET to monitor Tat-substrate-CFP with Tat-machinery-BiFC interactions. (A) When CFP or YFP are expressed, excitation of the protein fluorophore (434nm or 513nm, respectively) results in fluorescence (477nm or 527nm, respectively). However if the CFP and YFP are within 10nm of one another, fluorescence resonance energy transfer (FRET) can occur, whereby the CFP is excited at 434nm, emission occurs at 477nm but the non-radiative energy is absorbed by YFP, resulting in excitation of the fluorophore, and emission at 527nm. (B) Schematic diagram of how TatB-Y1 and TatC-Y2 when coexpressed reconstitute fluorescent YFP by BiFC. When ssDmsA-CFP is then expressed, the ssDmsA signal peptide directs the CFP to the TatBC complex, bringing the CFP into proximity and a subsequent YFP-BiFC-FRET signal can be generated.

CFP signal will occur demonstrating that FRET is occurring (Figure 5.1 A).

Utilizing the YFP-BiFC assay that we have developed to monitor protein interactions *in vivo* in the *E. coli* Tat pathway, we wished to additionally complex CFP

with another Tat pathway component to determine if a FRET signal could be detected between the protein complexes. Published experimental evidence has demonstrated that two *E. coli* Tat pathway proteins, NarJ and NarG, can be detected using CFP-YFP FRET (164) and the premise of CFP–YFP-BiFC–FRET has been demonstrated in eukaryotic cells (165,166). From these two lines of evidence, we are confident that a FRET signal can be generated *in vivo* in the Tat pathway.

5.3 Targeting interactions between Tat pathway machinery and protein substrates for FRET interrogation

From the YFP-BiFC analysis of the Tat pathway, many of the components used to detect interactions between proteins have been developed and described in Chapter 1. Full length CFP was fused C-terminally to either ssDmsA or TatC, generating two additional protein chimeras, ssDmsA-CFP and TatC-CFP, which allow for FRET analysis within the Tat pathway. For a FRET signal to occur the level of donor protein (CFP) has to be expressed at a lower level than the acceptor protein (YFP) because with a higher ratio of acceptor to donor protein the probability of energy transfer to the acceptor protein is increased and results in a higher signal to noise ration. We therefore placed the two CFP protein chimeras into a low copy arabinose inducible vector pBAD33 while the YFP-BiFC assay plasmids were left unaltered.

Two Tat pathway interactions that gave an appreciable YFP-BiFC signal were TatB-Y1/TatC-Y2 and ssDmsA-Y1/DmsD-Y2, so we added a third plasmid for FRET analysis to both of these systems resulting in TatB-Y1/TatC-Y2/ssDmsA-CFP and ssDmsA-Y1/DmsD-Y2/TatC-CFP (Figure 5.1 B and Figure 5.2, respectively). Since FRET is dependent on the relative signal ratio between the YFP to CFP, our negative control was CFP expressed from pBAD33 with both of the YFP-BiFC components

(e.g. TatB-Y1/TatC-Y2/CFP). CFP expression would result in a distribution of fluorescent protein evenly throughout the cytoplasm of *E. coli* and would not localize to the cell poles nor interact with the Tat pathway proteins.

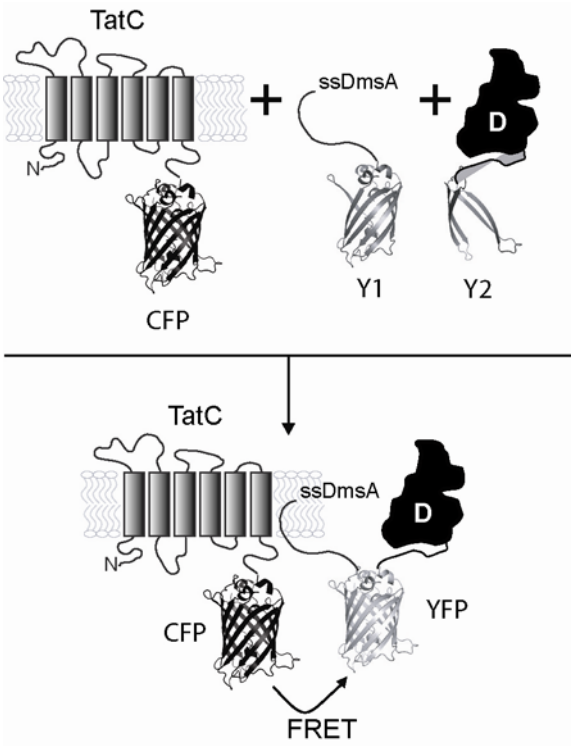

Figure 5.2 Applying YFP-BiFC-FRET to monitor TatC-CFP with YFP-BiFC ssDmsA-Y1/DmsD-Y2. (A) Coexpression of TatC-CFP (C-terminal fusion) in TG1 Δ*tatC* cells with YFP-BiFC ssDmsA-Y1/DmsD-Y2. The ssDmsA-Y1/DmsD-Y2 complex is targeted to TatC-CFP, therefore allowing for the interaction of ssDmsA with TatC-CFP, bringing YFP into proximity with CFP permitting FRET.

5.4 Preliminary experiments demonstrating YFP-BiFC FRET

The CFP–YFP-BiFC–FRET plasmids and the respective negative control CFP plasmid were transformed into Δ*tatC* TG1 cells. After induction of protein expression, a FRET signal was determined by taking the ratio of the YFP to CFP fluorescent signal intensity for both the FRET interacting plasmid combinations and for the FRET negative control. This ratio was then plotted against the relative fluorescence of the acceptor fluorophore (YFP) and a FRET signal could be determined from this data.

We were able to detect a FRET signal for the plasmid combinations ssDmsA-Y1/DmsD-Y2/TatC-CFP that had an apparent intensity that was 1.1–1.3 times greater than that for negative control (Figure 5.3). The lack of a FRET signal for the TatB-

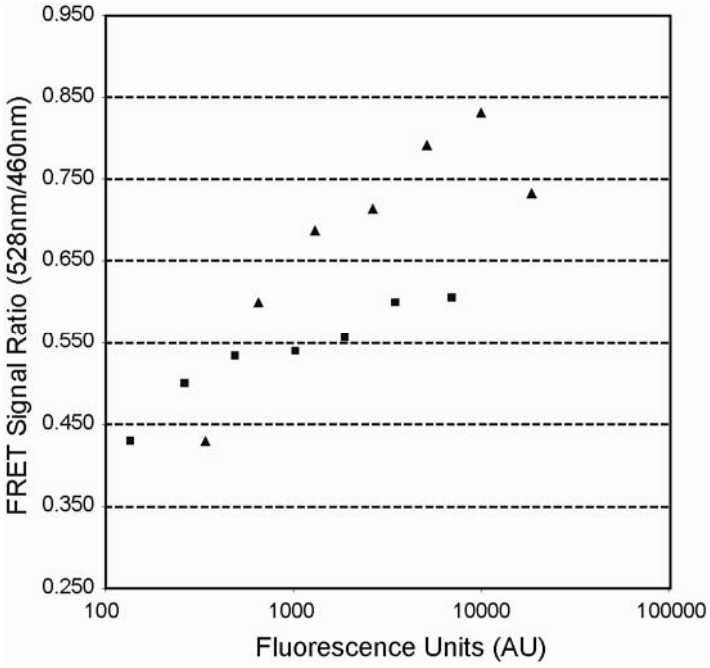

Figure 5.3 YFP-BiFC-FRET between TatC-CFP and ssDmsA-Y1/DmsD-Y2. Coexpression of either pBAD33Cm-TatC-CFP/ssDmsA-Y1/DmsD-Y2 (▲) or pBAD33Cm-CFP/ssDmsA-Y1/DmsD-Y2 (■) in Δ*tatC* TG1 cells. The FRET signal ratio observed for the pBAD33Cm-TatC-CFP/ssDmsA-Y1/DmsD-Y2 versus the pBAD33Cm-CFP/ssDmsA-Y1/DmsD-Y2 complex is plotted against the fluorescence units (AU) of YFP in each sample.

Y1/TatC-Y2/ssDmsA-CFP could be the result of numerous factors, a) Δ*tatC* TG1 cells actively exporting ssDmsA-CFP into the periplasm, b) the ratio between the acceptor (YFP) to donor (CFP) protein is incorrect and c) the distance between the two protein pairs is beyond the maximum 10nm, preventing FRET from occurring. Future experiments on optimizing the FRET conditions and addressing the factors described for the lack of a FRET signal will potentially clarify these results.

We have demonstrated for the first time that an YFP-BiFC-FRET signal can be detected in the *E. coli* Tat pathway. By using a combinatorial approach, proteins associated with the Tat pathway were expressed *in vivo* that first interacted to generate

a YFP-BiFC signal (e.g. ssDmsA-Y1/DmsD-Y2) and then interacted with the Tat translocon protein TatC-CFP. The cells were then lysed and *in vitro* experiments were performed to demonstrate that a FRET signal could be detected between the YFP-BiFC complex and the Tat-CFP translocon illustrating that the fluorescent proteins are within the Föster radius that allows for FRET to occur.

5.5 Discussion

With the development of two fluorescent *in vivo* techniques, BiFC and FRET, protein-protein interactions can be both spatially and temporally monitored in their native environments. We have demonstrated that YFP-BiFC can be applied to monitor protein interactions throughout the entirety of *E. coli* Tat pathway; however YFP-BiFC is limited in its ability to only report on two protein interactions while Tat transport involves numerous proteins that act in complex. Applying FRET alleviates this problem of YFP-BiFC by the addition of a second fluorescent protein (CFP) which in turn allows for the determination of the spatial organization of the proteins and also maintains the *in vivo* detection abilities of these assays.

We have demonstrated for the first time that a CFP–YFP-BiFC–FRET assay can be developed for the combinatorial *in vivo* and *in vitro* monitoring of protein complexes in the *E. coli* Tat pathway. Using two previously validated YFP-BiFC proteins (e.g. ssDmsA-Y1/DmsD-Y2) and a FRET protein partner (e.g. C-terminally tagged TatC-CFP) we were able to detect a FRET signal that was 1.1-1.3 fold over background. This evidence demonstrates that a combination of *in vivo* fluorescent proteins can be used to detect interactions between numerous proteins and that the distance between the Tat translocon machinery component TatC and the ssDmsA-Y1/DmsD-Y2 complex is within the Föster distance to allow for a FRET signal to be generated.

To further investigate how Tat machinery components assemble and interact, the application of a CFP-BiFC–YFP-BiFC–FRET assay has been envisioned. Evidence from the YFP-BiFC characterization of the Tat pathway demonstrated that numerous machinery proteins interact with one another (e.g. TatB-Y1/TatC-Y2) and with substrates directed for export (e.g. ssDmsA-Y1/TatC-Y2). By applying two different fluorescent BiFC assays, we hypothesize that we can capture two protein interactions with CFP-BiFC (e.g. TatB-C1/TatC-C2) and two protein interactions with YFP-BiFC (e.g. ssDmsA-Y1/DmsD-Y2) and then have the two BiFC complexes interact to generate a FRET signal. Optimization of the expression levels of the BiFC complexes will need to be determined along with the induction conditions for protein production because YFP and CFP BiFC complexes can cross associate (e.g. ssDmsA-Y1/TatC-C2) and generate aberrant complexes that would limit the detectable FRET signal.

Overall, we have demonstrated for the first time that a CFP–YFP-BiFC–FRET signal can be detected *in vivo* in the Tat pathway of *E. coli* and that this assay can be expanded into a CFP-BiFC–YFP-BiFC–FRET technique that can monitor how the Tat pathway functions as a complex of proteins.

5.6 Materials and Methods

5.6.1 Bacterial strains, growth and induction conditions.

Bacterial strains and growth conditions used in this study are described in Chapter 1.8. Induction of protein expression was done in a two step manner. After cells were grown to an OD_{600}~0.5 at 37ºC/200 rpm, 1mM IPTG was added and cells were transferred to RT/200 rpm and allowed to grow for 6 hours for expression of ssDmsA-Y1/DmsD-Y2 for YFP-BiFC. After 6 hours, 0.2% L-arabinose (final concentration) was added for induction of Tat-CFP expression from pBAD33Cm

TatC-CFP and cells were grown for another 10 hours at RT/200 rpm.

5.6.2 Construction of plasmids.

All plasmids used in this study are previously described in Chapter 1.8, with the exception of the plasmids listed below in Table 5.1. Plasmid pCyan (pBAD33Cm-CFP) encoding Cyan fluorescent protein (CFP) was provided as a gift from Dr. Patrick S. Daugherty (58). PCR amplification of CFP was preformed using forward primer MD2984 5' – CTC ATG GGT ACC GAA TTC GCG TAG CAT TGC GAC CTC TAA AGG TGA AGA ATT ATT CGG CGG – 3' and reverse primer MD2987 5' - CTC ATG ATC GAT CTT TAT TAT TTG TAC AAT TCA TCC ATA CCA TGG GTA ATA CCA GCA GCA – 3'. The PCR product was digested with *Kpn*I and *Cla*I and inserted into similarly digested pssDmsA-Y1 or pTatC-Y2 Tet^R, whereby the Y1 or Y2 fragments were replaced with CFP, respectively; generating pssDmsA-CFP and pTatC-CFP Tet^R. The TatC-CFP construct was then PCR amplified using forward primer MD3054 5' – CTC ATG GAG CTC AGG AGG AAT TCA CCA TGT CTG TAG AAG ATA CTC AAC CGC TTA TCA CGC – 3' and reverse primer MD3041 5' – CTC ATG GTC GAC TTA TTA TTT GTA CAA TTC ATC CAT ACC ATG GGT AAT ACC AGC AGC AGT – 3'. This PCR product was then digested with *Sac*I and *Sal*I and cloned into similarly digested pBAD33Cm creating the final plasmid pBAD33Cm-TatC-CFP.

5.6.3 FRET measurements and Data analysis

Cell density between samples pBAD33Cm-CFP/pssDmsA-Y1/pDmsD-Y2 and pBAD33Cm-TatC-CFP/pssDmsA-Y1/pDmsD-Y2 was normalized by OD_{600} readings. Cells (4-5mL) were centrifuged for 1 minute at 16,000 x g, 4ºC and subsequently treated with 400µL of Bug Buster Master Mix for 20 minutes at RT/200 rpm resulting in cell lysis. Cell lysate was then centrifuged for 20 minutes, 16,000 x g, at 4ºC and the supernatant (containing the soluble and membrane fractions) was collected.

Table 5.1 Plasmids used in this study.

Plasmid	Description	Reference
pBAD33Cm-CFP	pBAD33Cm expressing CFP from Nguyen *et. al.*	(58)
pssDmsA-CFP	The gene for CFP was inserted C-terminal to pssDmsA-Y1 replacing the Y1 fragment .	This work
pTatC-CFP Tet^R	The gene for CFP was inserted C-terminal to pTatC-Y2 Tet^R replacing the Y2 fragment.	This work
pBAD33Cm-ssDmsA-CFP	ssDmsA-CFP from pssDmsA-CFP was cloned into pBAD33Cm	This work
pBAD33Cm-TatC-CFP	TatC-CFP from pTatC-CFP Tet^R cloned into pBAD33Cm	This work

200μL of lysate was applied to the first column of a Costar flat bottom, high binding, and solid black 96 well plate. 100μL of lysate was serially diluted in 100μL of 1x PBS, resulting in a range of concentrations from 10^0 to 10^{-11}.

The 96 well plate was then placed into a Biotek Synergy HT plate reader and fluorescence measurements were determined using the following filter set combinations: CFP fluorescence (excitation 400/30, emission 460/40), YFP fluorescence (excitation 485/20, emission 528/20), and FRET (excitation 400/30, emission 528/20). After fluorescence data was acquired, the ratio between the FRET signal and the CFP signal was taken for dilutions 10^0 to 10^{-5}. This value is the FRET signal ratio (Figure 5.3) and it is plotted against the YFP signal obtained for each of the samples.

REFERENCES

1. Catherine S. Chan, T. M. L. W., Limei Chang, Charles M. Stevens, Mathew L. Workentine, Haiming Li, Ying Wei, Mary J. Ondrechen, Mark Paetzel, and Raymond J. Turner. (2008) *Biochemistry* **47**, 2749-2759
2. Yang Qiu, R. Z., T. Andrew Binkowski, Valentina Tereshko, Andrzej Joachimiak, Anthony Kossiakoff. (2008) *Proteins: Structure, Function, and Bioinformatics* **71**(2), 525-533
3. Walsh, S. T. R., Cheng, H., Bryson, J. W., Roder, H., and DeGrado, W. F. (1999) *Proceedings of the National Academy of Sciences of the United States of America* **96**(10), 5486
4. James W. Bryson, J. R. D., Tracy M. Handel, William F. Degrado. (1998) *Protein Science* **7**(6), 1404-1414
5. Philip A. Lee, D. T.-E., George Georgiou. (2006) *Annu Rev Microbiol.* **60**, 373-395.
6. DeLano, W. L. (2002) The PyMOL Molecular Graphics System. In.
7. Cline K, E. W., Theg SM. (1992) *J Biol Chem.* **267**(2688-96)
8. Mould RM, S. J., Robinson C. (1991) *J Biol Chem.* **266**, 17286-17289
9. Pugsley, A. P. (1993) *Microbiology and Molecular Biology Reviews* **57**(1), 50
10. Mori, H., and Ito, K. (2001) *Trends in Microbiology* **9**(10), 494
11. Driessen, A. J. M., Fekkes, P., and van der Wolk, J. P. W. (1998) *Current Opinion in Microbiology* **1**(2), 216
12. Driessen, A. J. M., Manting, E. H., and van der Does, C. (2001) *Nat Struct Mol Biol* **8**(6), 492
13. Wickner W, S. R. (2005) *Science.* **310**, 1452-1456.
14. Schatz, G., and Dobberstein, B. (1996) *Science* **271**(5255), 1519
15. Stuart, R. A., and Neupert, W. (2000) *Nature* **406**(6796), 575

16. Palmer, T., Sargent, F., Berks, B.C. (2005) *Treds in Microbiology* **13**(4), 175-180

17. Dilks K, R. R., Hartmann E, Pohlschröder M. (2003) *J Bacteriol.* **185**, 1478-1483.

18. Yen, M.-R., Tseng, Y.-H., Nguyen, E., Wu, L.-F., and Saier, M. (2002) *Archives of Microbiology* **177**(6), 441

19. Lee PA, T.-E. D., Georgiou G. (2006) *Annu Rev Microbiol.* **60**, 373-395.

20. Dilks K, R. R., Hartmann E, Pohlschröder M. (2003) *J Bacteriol.* **185**, 1478-1483.

21. Ochsner, U. A., Snyder, A., Vasil, A. I., and Vasil, M. L. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**(12), 8312

22. Voulhoux R, B. G., Ize B, Vasil ML, Lazdunski A, Wu LF, Filloux A. (2001) *EMBO J.* **20**, 6735-6741.

23. Lavander M, E. S., Bröms JE, Forsberg A. (2006) *Infect Immun.* **74**, 1768-1776

24. Saint-Joanis, B., Demangel, C., Jackson, M., Brodin, P., Marsollier, L., Boshoff, H., and Cole, S. T. (2006) *The Journal of Bacteriology* **188**(18), 6669

25. Cole, S. T. (2002) *European Respiratory Journal* **20**(36_suppl), 78S

26. Singh A, M. D., Kumar A, Steyn AJ. (2006) *Proc Natl Acad Sci.* **103**, 11346-11351.

27. Fernandes, P. (2006) *Nat Biotech* **24**(12), 1497

28. Jeong, K. J., Kawarasaki, Y., Gam, J., Harvey, B. R., Iverson, B. L., and Georgiou, G. (2004) *Journal of Molecular Biology* **341**(4), 901

29. Jack, R. L., Sargent, F., Berks, B. C., Sawers, G., and Palmer, T. (2001) *The Journal of Bacteriology* **183**(5), 1801

30. Stanley, N. R., Findlay, K., Berks, B. C., and Palmer, T. (2001) *The Journal of Bacteriology* **183**(1), 139

31. Lee, P. A., Buchanan, G., Stanley, N. R., Berks, B. C., and Palmer, T. (2002) *The Journal of Bacteriology* **184**(21), 5871

32. Philip A. Lee, G. L. O., Grant Buchanan, Nicholas P. Green, Peter J. Bond, Claire Punginelli, Rachel L. Jack, Mark S. P. Sansom, Ben C. Berks, and Tracy Palmer. (2006) *The Journal of Biological Chemistry* **281**, 34072-34085

33. Natascha Blaudeck, P. K., Matthias Muller, Georg A. Sprenger, and Roland Freudl. (2005) *The Journal of Biological Chemistry* **280**, 3426-3432

34. Matthew G. Hicks, P. A. L., George Georgious, Ben C. Berks, and Tracy Palmer. (2005) *Journal of Bacteriology* **187**, 2920-2925

35. Christopher A. McDevitt, M. G. H., Tracy Palmer, and Ben C. Berks. (2005) *Biochemical and Biophysical Research Communications* **329**, 693-698

36. Claire M. L. Barrett, D. M., and Colin Robinson. (2005) *Journal of Molecular Biology* **347**, 453-463

37. Claire M. L. Barrett, C. R. (2005) *FEBS Journal* **272**(9), 2261-2275

38. Strauch, E.-M., and Georgiou, G. (2007) *Journal of Molecular Biology* **374**(2), 283

39. Holzapfel E, E. G., Alami M, Barrett CM, Buchanan G, Lüke I, Betton JM, Robinson C, Palmer T, Moser M, Müller M. (2007) *Biochemistry.* **46**, 2892-2898.

40. Bolhuis, A., Mathers, J. E., Thomas, J. D., Barrett, C. M. L., and Robinson, C. (2001) *Journal of Biological Chemistry* **276**(23), 20213

41. Porcelli, I., de Leeuw, E., Wallis, R., van den Brink-van der Laan, E., de Kruijff, B., Wallace, B. A., Palmer, T., and Berks, B. C. (2002) *Biochemistry* **41**(46), 13690

42. de Leeuw, E., Granjon, T., Porcelli, I., Alami, M., Carr, S. B., Müller, M., Sargent, F., Palmer, T., and Berks, B. C. (2002) *Journal of Molecular Biology* **322**(5), 1135

43. Gohlke, U., Pullan, L., McDevitt, C. A., Porcelli, I., de Leeuw, E., Palmer, T., Saibil, H. R., and Berks, B. C. (2005) *Proc Natl Acad Sci U S A* **102**(30), 10482-10486

44. Cline, K., and McCaffery, M. (2007) *EMBO J* **26**(13), 3039-3049

45. Sargent, F., Stanley, N. R., Berks, B. C., and Palmer, T. (1999) *Journal of Biological Chemistry* **274**(51), 36073

46. Alami, M., Luke, I., Deitermann, S., Eisner, G., Koch, H. G., Brunner, J., and Muller, M. (2003) *Mol Cell* **12**(4), 937-946

47. Ivan J. Oresnik, C. L. L., and Raymond J. Turner. (2001) *Molecular Microbiology* **40**(2), 323-331

48. Hatzixanthis K, C. T., Oubrie A, Richardson DJ, Turner RJ, Sargent F. (2005) *Proc Natl Acad Sci.* **102**, 8460-8465.

49. Winstone TL, W. M., Sarfo KJ, Binding AJ, Haslam BD, Turner RJ. (2006) *Arch Biochem Biophys.* **455**, 89-97.

50. Genest, O., Seduk, F., Ilbert, M., Méjean, V., and Iobbi-Nivol, C. (2006) *Biochemical and Biophysical Research Communications* **339**(3), 991

51. Grant, B., Julien, M., Sander, B. N., David, J. R., Tracy, P., and Frank, S. (2008) *FEBS letters* **582**(29), 3979

52. Rodrigue A, C. A., Beck K, Müller M, Wu LF. (1999) *J Biol Chem* **274**(19), 13223-13228

53. Waraho, D., and DeLisa, M. P. (2009) *Proceedings of the National Academy of Sciences* **106**(10), 3692

54. Berks, B. C. (1996) *Mol Microbiol.* **22**(3), 393-404

55. Ribnicky, B., Van Blarcom, T., and Georgiou, G. (2007) *Journal of Molecular Biology* **369**(3), 631

56. Timothy L. Yahr, W. T. W. (2001) *European Molecular Biology Organization* **20**(10), 2472-2479

57. Oresnik IJ, L. C., Turner RJ. (2001) *Mol Microbiol.* **40**, 323-331.

58. Nguyen, A. W., and Daugherty, P. S. (2005) *Nat Biotech* **23**(3), 355

59. Tsien, R. Y. (1998) *Annual Review of Biochemistry* **67**(1), 509

60. Hu CD, C. Y., Kerppola TK. (2002) *Mol Cell. 2002* **9**, 789-798.

61. Kerppola, T. (2006) *Nat Rev Mol Cell Biol.* **7**, 449-456.

62. Beat Nyfeler, S. W. M., and Hans-Peter Hauri. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6350-6355

63. Karimova G, P. J., Ullmann A, Ladant D. (1998) *Proc Natl Acad Sci* **95**(5752-6)

64. Pelletier JN, C.-V. F., Michnick SW. (1998) *Proc Natl Acad Sci* **95**, 12141-12146

65. Rossi F, C. C., Blau HM. (1997) *Proc Natl Acad Sci* **94**, 8405-8410.

66. Thomas J. Magliery, C. G. M. W., Weilan Pan, Dennis Mishler, Indraneel Ghosh, Andrew D. Hamilton, and Lynne Regan. (2005) *Journal of the American Chemical Society* **127**, 146-157

67. Ghosh, I., Hamilton, A. D., Regan, L. (2000) *J. Am. Chem. Soc.* **122**, 5658-5659.

68. Stephen W. Michnick, P. H. E., Emily N. Manderson, Ingrid Remy, and Eduard Stefan. (2007) *Nature Reviews Drug Discovery* **6**, 569-582

69. Hu, C.-D., and Kerppola, T. K. (2003) *Nat Biotech* **21**(5), 539

70. Magliery, T. J., Wilson, C. G. M., Pan, W., Mishler, D., Ghosh, I., Hamilton, A. D., and Regan, L. (2005) *Journal of the American Chemical Society* **127**(1),

146

71. Ray N, O. J., Turner RJ, Robinson C. (2003) *FEBS Lett.* **534**, 156-160.

72. Gennis, R. B. (1996) Cellular and Molecular Biology. In. *American Society of Microbiology*, 2nd Edition. Ed.

73. Nicola Ray, J. O., Raymond J. Turner, and Colin Robinson. (2003) *Federation of European Biochemical Societies* **534**, 156-160

74. Rachael L. Jack, G. B., Alexandra Dubini, Kostas Hatzixanthis, Tracy Palmer, and Frank Sargent. (2004) *The European Molecular Biology Organization Journal* **23**(20), 3962-3972

75. Jack, R. L., Dubini, A., Palmer, T., and Sargent, F. (2005) *Biochemical Society Transactions* **33**(Pt 1), 105

76. Jach, G., Pesch, M., Richter, K., Frings, S., and Uhrig, J. F. (2006) *Nat Meth* **3**(8), 597

77. Strongin, D. E., Bevis, B., Khuong, N., Downing, M. E., Strack, R. L., Sundaram, K., Glick, B. S., and Keenan, R. J. (2007) *Protein Engineering Design and Selection* **20**(11), 525

78. Agnes Rodrigue, A. C., Konstanze Beck, Matthias Muller, and Long-Fei Wu. (1999) *The Journal of Biological Chemistry* **274**(19), 13223-13228

79. A. Wali Karzai, E. D. R., and Robert T. Sauer. (2000) *Nature Structural Biology* **7**, 449-455

80. DeLisa, M. P., Samuelson, P., Palmer, T., Georgiou, G. (2002) *J. Biol. Chem.* **277**(33), 29825-29831

81. Bronstein, P. A., Marrichi, M., Cartinhour, S., Schneider, D. J., and DeLisa, M. P. (2005) *The Journal of Bacteriology* **187**(24), 8450

82. Frank Sargent, B. C. B., and Tracy Palmer. (2002) *Archives of Microbiology* **178**, 77-84

83. Cristóbal S, d. G. J., Nielsen H, von Heijne G. (1999) *EMBO J.* **18**, 2982-2990.

84. Graubner, W., Schierhorn, A., and Bruser, T. (2007) *Journal of Biological Chemistry* **282**(10), 7116

85. Pérez-Rodríguez, R., Fisher, A. C., Perlmutter, J. D., Hicks, M. G., Chanal, A., Santini, C.-L., Wu, L.-F., Palmer, T., and DeLisa, M. P. (2007) *Journal of Molecular Biology* **367**(3), 715

86. Catherine S. Chan, J. M. H., Matthew L. Workentine, and Raymond J. Turner. (2006) *Biochemical and Biophysical Research Communications* **343**, 244-251

87. Vergnes, A., Pommier, J., Toci, R., Blasco, F., Giordano, G., and Magalon, A. (2006) *Journal of Biological Chemistry* **281**(4), 2170

88. Pommier, J., Mejean, V., Giordano, G., and Iobbi-Nivol, C. (1998) *Journal of Biological Chemistry* **273**(26), 16615

89. Oates, J., Barrett, C. M. L., Barnett, J. P., Byrne, K. G., Bolhuis, A., and Robinson, C. (2005) *Journal of Molecular Biology* **346**(1), 295

90. Richter, S., and Bruser, T. (2005) *Journal of Biological Chemistry* **280**(52), 42723

91. Cline, K., and Mori, H. (2001) *The Journal of Cell Biology* **154**(4), 719

92. Leake, M. C., Greene, N. P., Godun, R. M., Granjon, T., Buchanan, G., Chen, S., Berry, R. M., Palmer, T., and Berks, B. C. (2008) *Proceedings of the National Academy of Sciences* **105**(40), 15376

93. Jana, B., Ute, L., and Thomas, B. s. (2007) *FEBS letters* **581**(21), 4085

94. George L. Orriss, M. J. T., Bérengère Ize, Frank Sargent, Susan M. Lea, Tracy Palmer, and Ben C. Berks. (2007) *FEBS letters* **581**(21), 4091

95. Berthelmann, F., Mehner, D., Richter, S., Lindenstrauss, U., Lunsdorf, H., Hause, G., and Bruser, T. (2008) *Journal of Biological Chemistry* **283**(37), 25281

96. Frank Sargent, E. G. B., Nicola R. Stanley, Margaret Wexler, Colin Robinson, Ben C. Berks, and Tracy Palmer. (1998) *The European Molecular Biology Organization Journal* **17**, 3640-3650

97. Felix Berthelmann, T. B. (2004) *Federation of European Biochemical Societies* **569**, 82-88

98. Christopher A. McDevitt, G. B. F. S. T. P. B. C. B. (2006) *FEBS Journal* **273**(24), 5656-5668

99. Alami M, T. D., Wu LF, Müller M. (2002) *J Biol Chem.* **277**, 20499-20503.

100. Hou, B., Frielingsdorf, S., and Klösgen, R. B. (2006) *Journal of Molecular Biology* **355**(5), 957

101. Buchanan, G., Sargent, F., Berks, B., and Palmer, T. (2001) *Archives of Microbiology* **177**(1), 107

102. Russell, J. S. a. D. W. (2001) *Molecular Cloning - A Laboratory Manual*, 3rd Edition Ed., Cold Spring Harbor Lab Press, Cold Spring Harbor, NY

103. Tomoya Baba, T. A., Miki HAsegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L. Wanner, and Hirotada Mori. (2006) *Molecular Systems Biology* **2**, 1-11

104. Datsenko, K. A., and Wanner, B. L. (2000) *Proc Natl Acad Sci U S A* **97**(12), 6640-6645

105. Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998) *Proc Natl Acad Sci U S A* **95**(10), 5752-5756

106. Chang-Deng Hu, Y. C., and Tom K. Kerppola. (2002) *Molecular Cell* **9**(4), 789-798

107. Guido Jach, M. P., Klaus Richter, Sabine Frings, and Joachim F. Uhrig. (2006) *Nat Meth* **3**(8), 597

108. Bronstein, P. A., Marrichi, M., Cartinhour, S., Schneider, D. J., and DeLisa, M.

P. (2005) *J Bacteriol* **187**(24), 8450-8461

109. Brown, T. A. (2002) *Genomes*, 2 Ed., John Wiley & Sons Inc., New York

110. Yuan, L., Kurek, I., English, J., and Keenan, R. (2005) *Microbiology and Molecular Biology Reviews* **69**(3), 373

111. Jozef Hanes, a. A. P. (1997) *Proceedings of the National Academy of Sciences of the United States of America* **94**(10), 4937

112. Smith, G. P. (1985) *Science* **228**(4705), 1315

113. Fisher, A. C., and DeLisa, M. P. (2009) *Journal of Molecular Biology* **385**(1), 299

114. Adam C. Fisher, W. K., and Matthew P. Delisa. (2006) *The Protein Society* **15**(3), 449-458

115. Deng, X. K., Nesbit, L. A., and Morrow, K. J., Jr. (2003) *Clinical and Vaccine Immunology* **10**(4), 587

116. Kontermann, R. E., and Müller, R. (1999) *Journal of Immunological Methods* **226**(1-2), 179

117. Boss, M. A., Kenten, J. H., Wood, C. R., and Emtage, J. S. (1984) *Nucleic Acids Research* **12**(9), 3791

118. de Wildt, R. M. T., Mundy, C. R., Gorick, B. D., and Tomlinson, I. M. (2000) *Nat Biotech* **18**(9), 989

119. Chan, C. S., Winstone, T. M., Chang, L., Stevens, C. M., Workentine, M. L., Li, H., Wei, Y., Ondrechen, M. J., Paetzel, M., and Turner, R. J. (2008) *Biochemistry* **47**(9), 2749-2759

120. Qiu, Y. (2008) *Proteins* **71**, 525-533

121. Chan, C. S., Chang, L., Rommens, K. L., and Turner, R. J. (2009) *The Journal of Bacteriology*, JB.00949-00908

122. Fang X, Z. W. (2008) *J Proteomics* **71**(3), 284-303

123. Roque A.C., L. C. R. (2008) *Methods Mol biol* **421**, 1-21

124. Melissa R. Junttila, S. S. T. S. J. K. J. W. (2005) *PROTEOMICS* **5**(5), 1199-1203

125. Michnick, S. W., Remy, I., Campbell-Valois, F.-X., Vallée-Bélisle, A., Pelletier, J. N., and Jeremy Thorner, S. D. E. a. J. N. A. (2000) Detection of protein-protein interactions by protein fragment complementation strategies. In. *Methods in Enzymology*, Academic Press

126. Pelletier, J. N., Arndt, K. M., Pluckthun, A., and Michnick, S. W. (1999) *Nat Biotech* **17**(7), 683

127. T. P. Hopp, K. S. P., V.L. Price, R.T. Libby, C.J. March, D.P. Cerretti, D.L. Urdal and P.J. Conlon. (1988) *Bio-Technology* **6**, 1204-1210

128. Röthlisberger, D., Pos, K. M., and Plückthun, A. (2004) *FEBS Letters* **564**(3), 340

129. Rasmussen, S. G. F., Choi, H.-J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R. P., Sanishvili, R., Fischetti, R. F., Schertler, G. F. X., Weis, W. I., and Kobilka, B. K. (2007) *Nature* **450**(7168), 383

130. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K., and Stevens, R. C. (2007) *Science* **318**(5854), 1258

131. Quido A.Valent, P. A. S., Stephen High, Jan-Willem L.de Gier, Gunnar von Heijne, Georg Lentzen, Wolfgang Wintermeyer, Bauke Oudega and Joen Luirink. (1998) *The European Molecular Biology Organization Journal* **17**(9), 2504-2512

132. Müller M, K. R. (2005) *Mol Membr Biol* **22**(1-2), 113-121

133. Cline, K. a. T., S.M. (2007) *The Sec and Tat protein translocation pathways in*

chloroplasts. In: The Enzymes, Molecular Machines Involved in Protein Transport across Cellular Membranes, Elsevier, San Diego, CA

134. DeLisa, M. P., Tullman, D., and Georgiou, G. (2003) *Proceedings of the National Academy of Sciences of the United States of America* **100**(10), 6115
135. Carsten Sanders, N. W. H. L. (2001) *Molecular Microbiology* **41**(1), 241-246
136. Gérard F, P. N., Wu LF. (2005) *J Bacteriol.* **187(6)**, 1945-1950.
137. Ize B, G. F., Zhang M, Chanal A, Voulhoux R, Palmer T, Filloux A, Wu LF. (2002) *J Mol Biol.* **317**, 327-335
138. Robinson, C., and Bolhuis, A. (2001) *Nat Rev Mol Cell Biol* **2**(5), 350
139. Kenneth Cline, M. M. (2007) *The European Molecular Biology Organization Journal* **26**, 3039-3049
140. Blundell, T. L., Jhoti, H., and Abell, C. (2002) *Nat Rev Drug Discov* **1**(1), 45
141. Blundell, T. L., Sibanda, B. L., Montalvão, R. W., Brewerton, S., Chelliah, V., Worth, C. L., Harmer, N. J., Davies, O., and Burke, D. (2006) *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**(1467), 413
142. Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) *Nature* **448**(7155), 775
143. Beasley, J. R., and Hecht, M. H. (1997) *Journal of Biological Chemistry* **272**(4), 2031
144. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. (1995) *Science* **270**(5238), 935
145. Choo, Y., Castellanos, A., García-Hernández, B., Sánchez-García, I., and Klug, A. (1997) *Journal of Molecular Biology* **273**(3), 525
146. Pomerantz, J. L., Wolfe, S. A., and Pabo, C. O. (1998) *Biochemistry* **37**(4), 965
147. Struthers, M. D., Cheng, R. P., and Imperiali, B. (1996) *Journal of the American Chemical Society* **118**(13), 3073

148. Dahiyat, B. I., and Mayo, S. L. (1997) *Science* **278**(5335), 82

149. Harbury, P. B., Kim, P. S., and Alber, T. (1994) *Nature* **371**(6492), 80

150. Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q., and Baker, D. (1997) *Nat Struct Mol Biol* **4**(10), 805

151. Kevin H. Mayo, E. I. H. P. (1996) *Protein Science* **5**(7), 1301-1315

152. Karyn T. O'Neil, W. F. D. (1990) *Science* **250**, 646-651

153. Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., DeGrado, W. F., and Eisenberg, D. (1993) *Science* **259**(5099), 1288

154. Wei, Y., Kim, S., Fela, D., Baum, J., and Hecht, M. H. (2003) *Proceedings of the National Academy of Sciences of the United States of America* **100**(23), 13270

155. Edman, P. (1950) *Acta Chemica Scandinavica* **4**, 283

156. Patterson, S. D. (1994) *Analytical Biochemistry* **221**(1), 1

157. Gilmour, D. S., and Lis, J. T. (1984) *Proceedings of the National Academy of Sciences of the United States of America* **81**(14), 4275

158. Dedon, P. C., Soults, J. A., Allis, C. D., and Gorovsky, M. A. (1991) *Molecular and Cellular Biology* **11**(3), 1729

159. Fields, S., and Song, O.-k. (1989) *Nature* **340**(6230), 245

160. Joung, J. K., Ramm, E. I., and Pabo, C. O. (2000) *Proceedings of the National Academy of Sciences of the United States of America* **97**(13), 7382

161. Förster, T. (1948) *Annalen der Physik* **437**(1-2), 55-75

162. Lakowicz, J. R. (1999) *Energy Transfer. In Principles of fluorescence spectroscopy.*, 2nd ed. Ed., Kluwer Academic/Plenum Publishers, New York

163. Takanishi, C. L., Bykova, E. A., Cheng, W., and Zheng, J. (2006) *Brain Research* **1091**(1), 132

164. Haiming Li, R. J. T. (2009) *Canadian Journal of Microbiology* **55**(2), 179-188

165. Shyu, Y. J., Suarez, C. D., and Hu, C.-D. (2008) *Proceedings of the National Academy of Sciences* **105**(1), 151

166. Shyu, Y. J., and Hu, C.-D. (2008) *Trends in Biotechnology* **26**(11), 622